Computer Networks

Experimental Analysis of Localization Algorithms with NLOS Pre-Mitigation

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Non Line-Of-Sight (NLOS); channel status identification; positioning
Corresponding Author:	Marco Martalo University of Cagliari: Universita degli Studi Di Cagliari Cagliari, Sardinia ITALY
First Author:	Marco Martalo
Order of Authors:	Marco Martalo
	Fabrizio Carpi
	Luca Davoli
	Antonio Cilfone
	Yingjie Yu
	Yi Wang
	Gianluigi Ferrari
Abstract:	In this paper, we investigate enhanced target localization with mitigation of Non-Line-Of-Sight (NLOS) wireless links. Our approach is based on the acquisition, at the receiver, of a sequence of consecutive measurements of the Received Signal Strength Indicator (RSSI) of the signals transmitted by reference nodes, denoted as anchors . The RSSI is available in commercial mobile devices, including Internet of Things (IoT) devices. After the estimation of the status of each target-anchor link, on the basis of relevant statistical features extracted from RSSI values, a NLOS link is “transformed” into an equivalent LOS one and the corresponding distance is then estimated. The estimated distances feed “agnostic” localization algorithms, operating as if all links were LOS. We experimentally assess the performance of our approach in indoor (IEEE 802.11-based) and outdoor (Long Term Evolution, LTEbased) scenarios, considering both geometric and Particle Swarm Optimization (PSO)-based localization algorithms. Our results show the effectiveness of the proposed NLOS mitigation strategy towards accurate localization. In particular, the considered classifier, based on RSSI statistical features, performs similarly to other schemes appeared in the literature. Our results show that a relatively high (even for IoT scenarios) localization accuracy can be achieved especially in the IEEE 802.11-based indoor case (with six anchors). Key words: Line-Of-Sight (LOS), Non-LOS (NLOS), channel status identification, IEEE 802.11, LTE, NLOS mitigation, positioning.
Suggested Reviewers:	Evsen Yanmaz Assistant Professor, Ozyegin University: Ozyegin Universitesi eyanmaz@alumni.cmu.edu
	Antonella Molinaro Associate Professor, University of Reggio Calabria: Universita degli Studi Mediterranea di Reggio Calabria antonella.molinaro@unirc.it
Opposed Reviewers:	

Experimental Analysis of Localization Algorithms with NLOS Pre-Mitigation

Fabrizio Carpi,[1] Marco Martalò,[2] Luca Davoli,[3] Antonio Cilfone,[4] Yingjie Yu,[5] Yi Wang,[5] and Gianluigi Ferrari[3]

1: New York University (NYU), NY, USA
2: Department of Electrical and Electronic Engineering, University of Cagliari, and CNIT, Cagliari, Italy
3: Internet of Things (IoT) Laboratory, Department of Engineering and Architecture, University of Parma, CNIT, and CINI, Parma, Italy
4: Tesmec Automation s.r.l., Fidenza, Italy
5: Huawei Technologies Co. Ltd., Shanghai, China
E-mail: fabrizio.carpi@nyu.edu, marco.martalo@unica.it,
{luca.davoli,gianluigi.ferrari}@unipr.it, antonio.cilfone@tesmec.com,
{yuyingjie1,yi.wang}@huawei.com[1]

Abstract

In this paper, we investigate enhanced target localization with mitigation of Non-Line-Of-Sight (NLOS) wireless links. Our approach is based on the acquisition, at the receiver, of a sequence of consecutive measurements of the Received Signal Strength Indicator (RSSI) of the signals transmitted by reference nodes, denoted as *anchors*. The RSSI is available in commercial mobile devices, including Internet of Things (IoT) devices. After the estimation of the status of each target-anchor link, on the basis of relevant statistical features extracted from RSSI values, a NLOS link is "transformed" into an equivalent LOS one and the corresponding distance is then estimated. The estimated distances feed "agnostic" localization algorithms, operating as if all

[1]F. Carpi, M. Martalò, and A. Cilfone were with the IoT Lab of the University of Parma when contributing to this work

links were LOS. We experimentally assess the performance of our approach in indoor (IEEE 802.11-based) and outdoor (Long Term Evolution, LTE-based) scenarios, considering both geometric and Particle Swarm Optimization (PSO)-based localization algorithms. Our results show the effectiveness of the proposed NLOS mitigation strategy towards accurate localization. In particular, the considered classifier, based on RSSI statistical features, performs similarly to other schemes appeared in the literature. Our results show that a relatively high (even for IoT scenarios) localization accuracy can be achieved especially in the IEEE 802.11-based indoor case (with six anchors).
Key words: Line-Of-Sight (LOS), Non-LOS (NLOS), channel status identification, IEEE 802.11, LTE, NLOS mitigation, positioning.

1. Introduction

User localization is a crucial requirement for modern networks (e.g., cellular networks), since it allows providers to offer enhanced location-based services [1]. The Internet of Things (IoT) will benefit from these services, as adding location information will limit human intervention [2].

Radio-based positioning relies on distance estimates between the target and a few reference (with known positions) nodes, denoted as *anchors*. Such estimates can be obtained from relevant parameters, depending on the considered radio technology, such as: Received Signal Strength Indicator (RSSI), Angle of Arrival (AoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA) [3]. In [4], an enhanced fingerprinting method is considered, by relying on crowdsourced data kept from smartphones. In [5], a few positioning techniques are analyzed considering various experimental IoT wireless tech-

nologies: Zigbee, Bluetooth Low Energy (BLE), and WiFi (2.4 GHz band). In [6], AoA fingerprinting is improved by leveraging the available Channel State Information (CSI). In [7], RSSI-based localization is improved by means of Machine Learning (ML)-based techniques and, in particular, using Deep Reinforcement Learning (DRL).

User and device localization has been already exploited in IoT-based applications by leveraging various technologies. In the presence of large networking scenarios, the number of acquired RSSI data may explode. In order to reduce the amount of acquired RSSI data, a compression method is proposed in [8]. Low-complexity data processing for positioning in resource-constrained devices is also addressed in [9]. In [10], low-complexity RSSI-based localization in WiFi networks is proposed. In [11], visible light communications are considered to achieve a centimeter level accuracy. Range-free methods can also be considered to limit the complexity of the localization system, as shown in [12].

Radio-based positioning algorithms are impaired by physical obstructions and interference present in the surrounding environment, especially in indoor scenarios. In particular, in the presence of Line-Of-Sight (LOS) communications between the target and the anchors, the reliability of the position estimate may be very high. On the other hand, in the presence of Non-Line-Of-Sight (NLOS) links, the reliability of the position estimate may drastically reduce [13]. Therefore, the problem of channel status (i.e., LOS/NLOS) identification is crucial in radio-based localization—see, for example, [14] and references therein. In [15, 16], the LOS/NLOS classification problem is tackled, from an experimental point of view, in indoor IEEE 802.11-based scenar-

ios. In [16], the authors propose a Deep Learning (DL)-based method in which RSSI and CSI data are jointly used for classification. In [17], DL is used for LOS/NLOS link classification in Ultra-WideBand (UWB) scenarios. In [15], we have proposed channel status identification based on thresholding of simple RSSI statistical features. The use of thresholding on simple RSSI statistical features for LOS/NLOS identification purposes is also considered in Radio Frequency IDentification (RFID) schemes. For instance, in [18] the authors propose a threshold-based method on the variance of RSSI and received phase values.

In general, the performance of a localization strategy can be improved by properly taking into account the presence of NLOS links. In [19], the authors discuss NLOS identification and mitigation in UWB scenarios: channel identification and mitigation algorithms are based on ML techniques fed by fine-grained features extracted from the received waveform (e.g., received energy, maximum amplitude, rise time, etc.) and acquired with extensive experimental measurement campaigns. A similar scheme is proposed in [20]. In [21], the authors propose identification and mitigation in indoor WiFi scenarios, by relying on fine-grained CSI at the physical layer. In [22], a UWB identification and mitigation approach based on a Convolutional Neural Network (CNN) architecture is proposed. This method leverages the availability of the Channel Impulse Response (CIR), which obviously provides more information on the channel status than simple RSSI. However, the use of RSSI is attractive in IoT scenarios with constrained nodes.

In this paper, we consider a static localization scenario and focus on mitigating NLOS ranging errors using RSSI measurements in IEEE 802.11 and

Long Term Evolution (LTE) scenarios. It is known that RSSI in NLOS links can be characterized by statistical features, such as skewness, kurtosis, and others—see, e.g., [15, 16]. We leverage the identification method proposed in [15] to define and implement a NLOS mitigation scheme for enhanced localization. In particular, we use the feature-based classifier in [15] as a first processing stage to detect the presence of NLOS communication links; then, we exploit a linear combination of the same features for mitigating NLOS ranging errors. Finally, after NLOS measurements are "corrected," a localization algorithm is applied to estimate the target position.

Unlike previous literature works, we do not rely on fine-grained CSI, but we exploit RSSI values, which can be easily obtained with Commercial Off-The-Shelf (COTS) devices, such as IoT devices and smartphones. More precisely, our proposed localization approach involves the following two steps.

- The *first* pre-processing step consists in extracting five statistical features from N consecutive RSSI measurements from each link with an anchor. These features are then compared to pre-determined thresholds to classify the status of the link [15]. Finally, these features are regressed to derive a common correction parameter for NLOS links, in order to mitigate the link distance error induced by NLOS effects and, thus, "transform" NLOS links into equivalent LOS ones. This approach is different from previous literature works, see, e.g., [23, 24], in which the correction is achieved by means of more sophisticated algorithms requiring an *a-priori* statistical characterization of LOS/NLOS communication channels.

- In the *second* step, all estimated link distances are given as input to an

"agnostic" localization algorithm, which operates as if all links were in LOS conditions. In other words, we investigate the suitability of existing localization algorithms to be integrated with the proposed NLOS link pre-processing stage. The performance of the improved localization algorithms is then analyzed.

An experimental IoT-oriented performance analysis of the proposed localization strategies is carried out in both indoor (IEEE 802.11) and outdoor (LTE) scenarios.[2] Our results show that, in the best cases, the positioning error is around (approximately) 30% and 60% of the maximum target-anchor distance in indoor and outdoor scenarios, respectively.

The rest of the paper is organized as follows. In Section 2, we introduce the system model. In Section 3, the proposed localization method with NLOS identification/mitigation is presented. Experimental results for IoT-oriented IEEE 802.11 and LTE systems are discussed in Section 4. Finally, concluding remarks are given in Section 5.

2. System Model

Let us consider a wireless scenario, in which a static[3] target node, at coordinates $\boldsymbol{u} = [x, y]^T$, receives packets from M transmitters, either Access Points (APs) or Base Transceiver Stations (BTSs), acting as anchors. The

[2]For LTE-based analysis, we rely on the use of a smartphone. However, the obtained results are also meaningful for 4G NarrowBand-IoT (NB-IoT) scenarios, as the same RSSI values can be exploited.

[3]Locating a mobile node is an interesting research direction, but goes beyond the scope of this paper.

known position of the i-th anchor is denoted as $\boldsymbol{s}_i = [x_i, y_i]^T$, $i \in \{1, \ldots, M\}$, where $[\cdot]^T$ is the transpose operator. The set of anchor nodes' coordinates can be organized in the following matrix:

$$\boldsymbol{S} = \begin{bmatrix} x_1 & x_2 & \ldots & x_M \\ y_1 & y_2 & \ldots & y_M \end{bmatrix} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_M]\,. \tag{1}$$

This scenario is meaningful for applications in which targets (people and/or objects) to be localized are moving on the $x - y$ plane, e.g., IoT tags moving on a given building floor. The extension to a three-dimensional case is straightforward, but goes beyond the scope of this paper.

The goal of a localization system is to derive an estimate of the target node's position, denoted[4] as $\hat{\boldsymbol{u}} = [\hat{x}, \hat{y}]^T$, given $\boldsymbol{S}$ and a set of measurements of the target-anchor links. The Euclidean norm of the i-th anchor coordinates, i.e., the distances of that anchor from the axes' origin, is defined as

$$k_i = \|\boldsymbol{s}_i\| = \sqrt{\boldsymbol{s}_i^T \boldsymbol{s}_i} \qquad i = 1, \ldots, M.$$

The anchors' norm vector is then $\boldsymbol{k} = [k_1, \ldots, k_M]^T$.

Let $\boldsymbol{d} = [d_1, d_2, \ldots, d_M]^T$ be the vector containing the (true) link distances between the target node and the M anchors, where the i-th distance can be written as

$$d_i = \|\boldsymbol{u} - \boldsymbol{s}_i\| = \sqrt{(\boldsymbol{u} - \boldsymbol{s}_i)^T (\boldsymbol{u} - \boldsymbol{s}_i)} \quad i = 1, \ldots, M. \tag{2}$$

[4]In the remainder of this paper, the symbol $\hat{\zeta}$ will denote an estimate of the quantity ζ, e.g., calculated from experimental measurements or inferred through proper signal processing.

Let $\hat{\boldsymbol{d}}$ denote the corresponding vector of distance estimates, in which the i-th term can be written as

$$\hat{d}_i = \|\hat{\boldsymbol{u}} - \boldsymbol{s}_i\| = \sqrt{(\hat{\boldsymbol{u}} - \boldsymbol{s}_i)^T(\hat{\boldsymbol{u}} - \boldsymbol{s}_i)} \quad i = 1, \ldots, M. \tag{3}$$

For instance, $\hat{d}_i$ is the i-th link distance estimate computed from the acquired data (e.g., RSSIs or ToAs).

The localization problem can be generally modeled as the following system of equations containing anchors' positions and range estimates:

$$\begin{cases} (\hat{x} - x_1)^2 + (\hat{y} - y_1)^2 = \hat{d}_1^2 \\ \qquad\qquad \vdots \\ (\hat{x} - x_M)^2 + (\hat{y} - y_M)^2 = \hat{d}_M^2. \end{cases} \tag{4}$$

In order to solve this system of equations in the unknowns $[\hat{x}, \hat{y}]$, different solutions will be proposed in Subsection 3.4.

While ToA-based localization algorithms work on the basis of estimated distances $\{\hat{d}_i\}$, TDoA-based algorithms rely on relative distance estimates with respect to a reference anchor. Assuming, for notational simplicity, that $\boldsymbol{s}_1$ is the anchor with the shortest estimated distance, then the vector of relative distances (with respect to $\boldsymbol{s}_1$) is defined as $\boldsymbol{\Delta} = [\Delta_1, \ldots, \Delta_M]^T$, in which the i-th term is

$$\Delta_i = d_i - d_1 \tag{5}$$

and its estimate is

$$\hat{\Delta}_i = \hat{d}_i - \hat{d}_1. \tag{6}$$

Note that $\Delta_1 = \hat{\Delta}_1 = 0$ by construction. The quantities $\{\hat{\Delta}_i\}_{i=1}^M$ can be interpreted as TDoA-based relative distance estimates with respect to the first anchor.

At each time instant, a received power measurement, e.g., the RSSI, is acquired. In order to perform target-(anchor i) channel status classification, N consecutive RSSI measurements are collected into the following "observation" vector:

$$\boldsymbol{z}_i^{(j)} = \left[z_i^{((j-1)N+1)}, z_i^{((j-1)N+2)}, \ldots, z_i^{(jN)} \right]$$

where j is a block (of N RSSI values) time index.

The target-(anchor i) communication channel is assumed to have the same binary status over the RSSI observation block j, and we refer to this status as

$$\ell_i^{(j)} = \begin{cases} 1 & \text{if the link is LOS} \\ 0 & \text{if the link is NLOS.} \end{cases} \tag{7}$$

We denote the collection of K consecutive observation vectors and corresponding true channel statuses as the dataset relative to the i-th anchor, i.e.:

$$\mathcal{D}_i = \left\{ \boldsymbol{z}_i^{(j)}, \ell_i^{(j)} \right\}_{j=1}^{K} \qquad i = 1, \ldots, M. \tag{8}$$

In other words, $\boldsymbol{z}_i^{(1)}$ contains the first N acquired RSSI measurements associated with the true channel status $\ell_i^{(1)}$, and so on for $i > 1$. Note that no *a-priori* information about the target-(anchor i) distance is contained in the dataset $\mathcal{D}_i$, $i = 1, \ldots, M$. We assume that each RSSI entry in the observation vector $\boldsymbol{z}_i^{(j)}$ is a realization of a random variable $Z_i^{(j)}$ whose statistical distribution depends on the LOS/NLOS conditions of the target-(anchor i) link during the j-th RSSI observation block, i.e., $\ell_i^{(j)}$. The dependence of $Z_i^{(j)}$ on $\ell_i^{(j)}$ would depend on the very specific propagation conditions of the associated link. We give up finding an accurate statistical characterization of $Z_i^{(j)}$ but, rather, extract simple statistical features. More precisely, the dataset

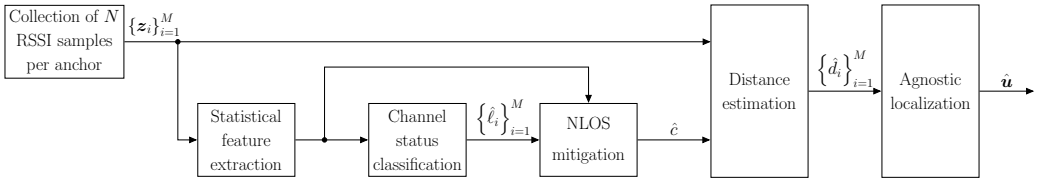


Figure 1: Block diagram of the proposed localization approach for a generic observation window.

$\mathcal{D}_i$ will be at the basis of the extraction of relevant statistical features of the received signal to be used to classify the status of the communication channel between the target and the i-th anchor.

3. Localization Approach

The block diagram associated with the proposed localization approach is shown in Figure 1. The localization method is divided in two steps. *First*, for each anchor we process the received signal (in terms of RSSI) to classify the LOS/NLOS status of the channel (Subsection 3.1) and to estimate the distance from the anchor (Subsection 3.2 for LOS and Subsection 3.3 for NLOS). *Then*, "agnostic" localization algorithms are used, taking as input the link distance estimates (for all anchors) obtained from the previous step (Subsection 3.4).

3.1. LOS/NLOS Classification

We now briefly recall the feature-based LOS/NLOS classification method proposed in [15]. Let us focus on a single target-anchor link over an ob-

servation window of N RSSI samples.[5] Given the observation vector $\boldsymbol{z} = [z^{(1)}, z^{(2)}, \ldots, z^{(N)}]$ and the corresponding random variable Z (as described at the end of Section 2), let us define the k-th order moment of the distribution of the random variable Z as

$$m_k = \frac{1}{N} \sum_{i=1}^{N} \left(z^{(i)} - \mu \right)^k \tag{9}$$

where

$$\mu = \frac{1}{N} \sum_{i=1}^{N} z^{(i)} \tag{10}$$

is the sample mean for the observation vector. The five considered statistical features are the following:

- standard deviation, defined as $\sigma = \sqrt{m_2}$;
- skewness, defined as $S = m_3/\sigma^3$;
- kurtosis, defined as $K = m_4/\sigma^4$;
- hyper-skewness, defined as $\mathscr{S} = m_5/\sigma^5$;
- Peak Probability (PP), defined as

$$\mathscr{P} = \Pr\left\{ z^{(i)} \in \left[\max_i f\left(z^{(i)}\right) - \epsilon, \max_i f\left(z^{(i)}\right) + \epsilon \right] \right\}$$

where ϵ is a "sufficiently" small non-negative value. In particular, if $\epsilon = 0$, $\mathscr{P}$ is equivalent to the relative frequency of the mode value within the observation vector $\boldsymbol{z}$. Moreover, skewness, kurtosis, and hyper-skewness

[5] The anchor subscript i and the observation window superscript j used in Section 2 are eliminated for notational simplicity.

can be referred to as third, fourth and fifth order standardized moments, respectively—the k-th order standardized moment is defined as m_k/σ^k ($k \geq 1$).

In correspondence to the observation vector $\boldsymbol{z}$, we define the feature dataset $\tilde{\mathcal{D}}$ as the set of statistical features extracted from $\boldsymbol{z}$ together with the true channel status ℓ, i.e.,

$$\tilde{\mathcal{D}} = \{\sigma, S, K, \mathscr{S}, \mathscr{P}, \ell\}\,. \tag{11}$$

Obviously, $\tilde{\mathcal{D}}$ depends on the (extended) dataset, defined in (8), associated with the link status and the RSSI observation block.

As outlined in [15], in LOS scenarios the direct path is characterized by a much higher received power than the reflected ones. Therefore, the RSSI distribution is expected to be peaky and *left-skewed*, and can then be well approximated by a Weibull distribution. On the other hand, in NLOS scenarios there may be several scattered and reflected paths with smaller received power. Consequently, the RSSI distribution is expected to be *symmetric* and less peaky, and can then be well approximated by a Gaussian distribution. Illustrative examples of RSSI distributions, for a given target location, in both LOS and NLOS link conditions are shown in Figure 2.

In [15], various classification algorithms are proposed: a Probability Mass Function (PMF)-based one; a Neural Network (NN)-based one; and a weighed threshold classifier. The first is shown to have very poor performance; on the other hand, the second is one of the possible Artificial Intelligence (AI)-based schemes suitable for such an application. In particular, a three-layer NN is considered with the following characteristics: (i) the input layer is composed by the five statistical features; (ii) the hidden layer is a 8-neuron fully con-

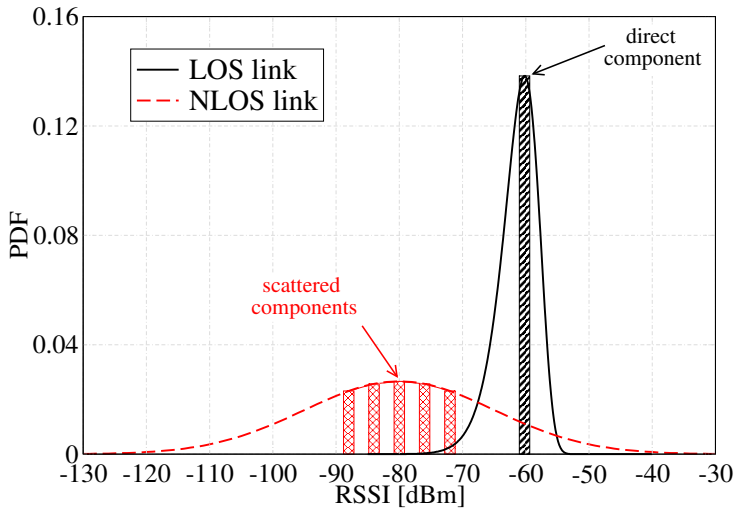


Figure 2: Illustrative examples of RSSI distributions, for a given target location, in both LOS and NLOS link conditions.

nected layer; and (iii) the output layer corresponds to the estimated channel condition. The activation functions for the hidden and output layers are *sigmoid* and *softmax* functions, respectively. Even if other AI algorithms can be applied (this goes beyond the scope of this paper), the NN can be considered as a reasonable state-of-the-art benchmark. In the following, we focus on the weighed threshold classifier, whose architecture is shown in Figure 3.

The key idea is to compare each statistical feature with properly chosen threshold values $\{\sigma^*, S^*, K^*, \mathscr{S}^*, \mathscr{P}^*\}$, in order to derive single-feature

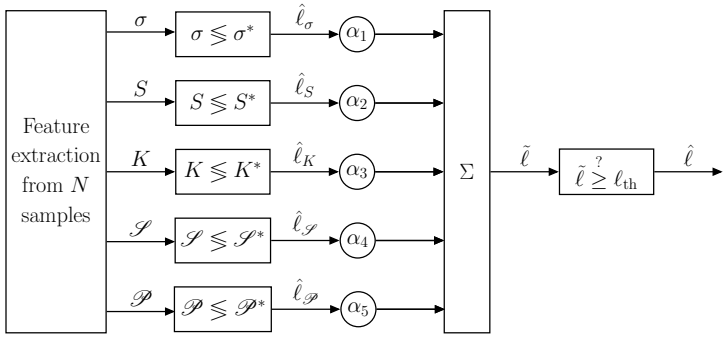


Figure 3: Block diagram of the majority logic threshold detector for NLOS classification.

channel estimates as follows:

$$\begin{aligned} \hat{\ell}_\sigma &= U(\sigma^* - \sigma) \\ \hat{\ell}_S &= U(S^* - S) \\ \hat{\ell}_K &= U(K - K^*) \\ \hat{\ell}_{\mathscr{S}} &= U(\mathscr{S}^* - \mathscr{S}) \\ \hat{\ell}_{\mathscr{P}} &= U(\mathscr{P} - \mathscr{P}^*) \end{aligned} \tag{12}$$

where $U(\cdot)$ is the unit-step function, i.e., it equals 1 for positive argument and 0 otherwise. The rationale behind (12) is that for LOS links, one expects smaller values of σ, S, and $\mathscr{S}$, and larger values of K and $\mathscr{P}$ for NLOS links. This is due to the fact that the direct component is dominant in LOS conditions, whereas the reflected paths provide a minor contribution. In particular, in LOS links, the odd-order statistical features S and $\mathscr{S}$ are expected to be negative (due to the left-skewed distribution). The chosen values of the feature thresholds depend on the considered wireless technology (namely, IEEE 802.11 and LTE) and corresponding network parameters (transmit power, RSSI resolution, etc.). However, such thresholds can be set

offline to achieve the desired classification accuracy, provided that the relevant parameters are known and/or fixed in advance. As shown in Section 4, in our numerical results we rely on measurements taken under different conditions and we identify a unique threshold that can achieve, on average, good performance.

The single-feature decisions can be collected into the vector $\hat{\boldsymbol{\ell}} = [\hat{\ell}_\sigma, \hat{\ell}_S, \hat{\ell}_K, \hat{\ell}_{\mathscr{S}}, \hat{\ell}_{\mathscr{P}}]^T$. At this point, each single decision is weighed by a proper coefficient $\alpha_i \in [0, 1]$, $i = 1, 2, \ldots, 5$, to take into account possibly different reliabilities of the single decisions. Let us denote as $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_5]^T$ the vector of weighing coefficients, with $\|\boldsymbol{\alpha}\|_1 = 1$. The final decision variable $\tilde{\ell} \in [0, 1]$ is

$$\tilde{\ell} = \boldsymbol{\alpha}^T \hat{\boldsymbol{\ell}} \tag{13}$$

and the channel status is finally estimated as

$$\hat{\ell} = U(\tilde{\ell} - \ell_{\text{th}}) \tag{14}$$

where $\ell_{\text{th}} \in [0, 1]$ is a proper threshold. Note that if $\boldsymbol{\alpha} = 0.2 \cdot \mathbf{1}_5$, being $\mathbf{1}_n$ the all-one vector of size n, this approach is equivalent to an unweighed majority logic threshold detector, i.e., all features have the same relevance. On the other hand, if one of the weights is equal to 1 (and, consequently, the others are 0), this rule is equivalent to a single-feature threshold detector for the feature corresponding to the unitary weight.

3.2. LOS Distance Estimation

Assuming the applicability of Friis formula, the estimated distance on a generic link, on the basis of the received power P_{R} (dimension: [dBm]), can be expressed as [25]

$$\hat{d} = d_0 \, 10^{\frac{P_0 - P_{\text{R}}}{10\,\beta}} \tag{15}$$

where P_0 is the received power (dimension: [dBm]) at the reference distance d_0 (dimension: [m]), and β is the path loss exponent (adimensional). Eq. (15) can be applied to transmissions over channels with LOS conditions, considering $\beta = 2$. Assume that the receiver can collect RSSI samples, where RSSI is the actual power level (dimension: [dBm]) seen by the receiver.[6] Considering the average RSSI (denoted as $\overline{\text{RSSI}}$) over the observation window of N samples, the estimated distance can be approximated as follows:

$$\hat{d} \simeq d_0 \, 10^{\frac{P_0 - \overline{\text{RSSI}}}{20}}. \tag{16}$$

Note that the use of the average RSSI is expedient to eliminate statistical fluctuations (especially in experimental scenarios).

3.3. NLOS Mitigation and Distance Estimation

As shown in Figure 2, for NLOS links a direct path between the transmitter and the receiver with dominant received power does not exist. Therefore, the electromagnetic signal travels along a longer path through reflections and/or refractions, thus reducing (with respect to the direct path) the received power. The application of (15) is critical, as the value of β would depend on the specific path. Therefore, the use of (15) with a fixed value of β (for instance, the use of (16)) may likely lead to a wrong distance estimate in NLOS conditions. To this end, we propose to mitigate the NLOS effect by deriving a more reliable link distance estimation strategy starting from (16), rather than adapting β in (15) link by link. In particular, we

[6] This assumption is typical for IoT devices. Moreover, in most cases the available RSSI values are quantized.

propose to transform a NLOS link into an "equivalent" LOS one, using a heuristic "universal" correction (independent of the specific link conditions), as will be described in the following.

Let us focus on the k-th observation window ($k = 1, \ldots, K$) of a generic target-anchor link. When the link is classified as NLOS (according to the strategy outlined in Subsection 3.1), the corresponding estimated distance $\hat{d}_{\text{nlos}}$ obtained with (16) may be heuristically corrected by means of a scalar coefficient as

$$\hat{d}^{(k)}_{\text{nlos}-\text{c}} = \frac{\hat{d}^{(k)}_{\text{nlos}}}{\hat{c}^{(k)}} \tag{17}$$

where $\hat{d}^{(k)}_{\text{nlos}-\text{c}}$ is the estimated distance after the correction and $\hat{c}^{(k)} > 1$ is a proper correction parameter, which quantifies the NLOS effect. In particular, the correction coefficient at the k-th time epoch can be empirically computed as a linear regression, with proper coefficients, of the statistical features extracted online from the current block of N RSSI samples. In other words,

$$\hat{c}^{(k)} = \boldsymbol{\gamma}^T \, \boldsymbol{f}^{(k)} \tag{18}$$

where $\boldsymbol{f}^{(k)} = [1, \sigma^{(k)}, S^{(k)}, K^{(k)}, \mathscr{S}^{(k)}, \mathscr{P}^{(k)}]^T$ is the vector of statistical features (besides the initial 1) and $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_5]^T$ is a fixed vector of regression weights for the features, which is kept constant, regardless of the specific link and observation window, in NLOS conditions. A pictorial description of the NLOS mitigation scheme is provided in Figure 4.

The regression weight vector $\boldsymbol{\gamma}$ can be computed offline during a system calibration phase as follows. Assume that the target is placed at known pre-defined positions, so that L estimated distances associated with NLOS

links are available[7] and collected in the vector $\hat{\boldsymbol{d}}_{\text{nlos}} = [\hat{d}^{(1)}_{\text{nlos}}, \ldots, \hat{d}^{(L)}_{\text{nlos}}]^T$. Assume also that the corresponding true distances are known and collected in the vector $\boldsymbol{d}_{\text{true}} = [d^{(1)}, \ldots, d^{(L)}]^T$. The statistical features over the entire observation window can be collected in the following $L \times 6$ matrix:

$$\boldsymbol{F} = \left[\boldsymbol{f}^{(1)^T}, \boldsymbol{f}^{(2)^T}, \ldots, \boldsymbol{f}^{(L)^T} \right]^T. \tag{19}$$

The optimal weight vector solves the following minimization problem:

$$\boldsymbol{\gamma} = \underset{\boldsymbol{a}=[a_0,\ldots,a_5]^T}{\arg\min} \left|\left| \boldsymbol{d}_{\text{true}} - \hat{\boldsymbol{d}}_{\text{nlos}} \odot \frac{1}{\boldsymbol{F}\,\boldsymbol{a}} \right|\right| \tag{20}$$

where $\odot$ denotes element-wise vector multiplication. In other words, the weight vector minimizes the error between the true distance and the corrected one. In order to make the calibration method accurate, the total number L of collected RSSI blocks must be sufficiently large. From a practical point of view, this corresponds to considering a sufficiently large number of NLOS links (sufficiently heterogeneous) with known distances and carry out data collection. The weight vector $\boldsymbol{\gamma}$ will then be used, during online operations, if a link is identified as NLOS. The identification of an efficient calibration dataset (in terms of set of L NLOS links) is an interesting research problem.

3.4. "Agnostic" Localization

Once the estimated distances are available from all the M anchors, an "agnostic" localization algorithm can be run to derive a final target position estimate. The algorithm is agnostic in the sense that it acts as if all links

[7] L consecutive N RSSI samples' blocks are collected sequentially.

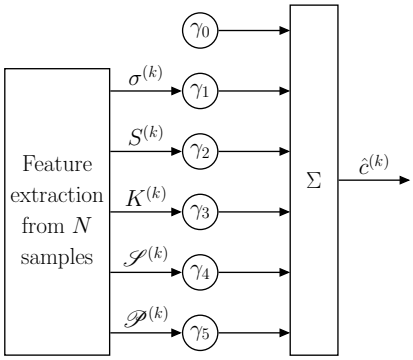


Figure 4: Block diagram of the proposed NLOS mitigation scheme.

were LOS. In fact, a NLOS link is transformed into an equivalent LOS one according to the steps discussed in Subsection 3.1 and Subsection 3.3.

In general, a localization algorithm solves a system of equations of the type shown in (4). As illustrative (but not exhaustive) examples of agnostic localization algorithms, we now briefly recall two geometric algorithms, namely Two-Stage Maximum-Likelihood (TSML) [26] and Plane Intersection (PI) [27], which will be used for localization purposes. In particular, TDoA-based implementation of TSML and PI are considered. Besides geometric solutions, Particle Swarm Optimization (PSO)-based solutions will also be considered. An exhaustive analysis of RSSI-based least squares lateration algorithms is presented in [28].

We remark that the localization accuracy of ToA-based algorithms, not shown here for lack of space, is lower than that of TDoA-based ones. Even if ToA-based processing may be more attractive from an implementation point of view, we focus on TDoA-based processing, which guarantees best performance. If TDoA-based processing is considered, inter-anchor synchronization is crucial. In IoT scenarios, this can be demanded to the infrastructure, i.e.,

to the cooperating anchors. The investigation of this aspect goes beyond the scope of this paper.

3.4.1. Two-Stage Maximum-Likelihood (TSML) [26]

The TSML algorithm resorts to a two-step approach and solves, in each step, a smaller system of equations (with respect to the starting one). Defining the unknown vector as $\boldsymbol{\phi}_1 = \left[\boldsymbol{u}^T, d_1\right]^T$, one obtains the following system of equations for the first step:

$$\boldsymbol{G}_1 \boldsymbol{\phi}_1 = \boldsymbol{h}_1 \tag{21}$$

where $\boldsymbol{G}_1$ is a $(M-1) \times 3$ matrix and $\boldsymbol{h}_1$ is a length-$(M-1)$ vector defined as follows:

$$\boldsymbol{G}_1 = -2 \begin{bmatrix} x_2 - x_1 & y_2 - y_1 & \hat{\Delta}_2 \\ \vdots & \vdots & \vdots \\ x_M - x_1 & y_M - y_1 & \hat{\Delta}_M \end{bmatrix} \tag{22}$$

$$\boldsymbol{h}_1 = \begin{bmatrix} k_1^2 - k_2^2 + \hat{\Delta}_2^2 \\ \vdots \\ k_1^2 - k_M^2 + \hat{\Delta}_M^2 \end{bmatrix}. \tag{23}$$

As $\boldsymbol{G}_1$ is not a square matrix, the solution $\hat{\boldsymbol{\phi}}_1$ can be obtained, by resorting to a Least Squares (LS) method [29], as follows:

$$\hat{\boldsymbol{\phi}}_1 = \left(\boldsymbol{G}_1^T \boldsymbol{G}_1\right)^{-1} \boldsymbol{G}_1^T \boldsymbol{h}_1. \tag{24}$$

At this point, since the third unknown in $\boldsymbol{\phi}_1$ (namely, d_1) depends on the other two (namely, $\boldsymbol{u}$), one has to solve a system of equations to eliminate this dependence. Denoting $\boldsymbol{\phi}_2 = (\boldsymbol{u} - \boldsymbol{s}_1)^2$, the following final system of

equations has to be solved:

$$\boldsymbol{G}_2\boldsymbol{\phi}_2 = \boldsymbol{h}_2 \tag{25}$$

where $\boldsymbol{G}_2$ is a 3×2 matrix, whereas $\boldsymbol{h}_2$ is a length-3 vector defined, respectively, as follows:

$$\boldsymbol{G}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \tag{26}$$

$$\boldsymbol{h}_2 = \left(\hat{\boldsymbol{\phi}}_1 - \left[\boldsymbol{s}_1^T, 0\right]^T\right)^2. \tag{27}$$

The LS solution is then

$$\hat{\boldsymbol{\phi}}_2 = \left(\boldsymbol{G}_2^T\boldsymbol{G}_2\right)^{-1}\boldsymbol{G}_2^T\boldsymbol{h}_2 \tag{28}$$

and the final position estimate is obtained combining $\hat{\boldsymbol{\phi}}_1$ and $\hat{\boldsymbol{\phi}}_2$ according to

$$\hat{\boldsymbol{u}} = \text{sign}\left(\breve{\boldsymbol{\phi}}_1 - \boldsymbol{a}_1\right) \odot \hat{\boldsymbol{\phi}}_2 + \boldsymbol{a}_1 \tag{29}$$

where $\text{sign}(\cdot)$ represents the sign operator and $\breve{\boldsymbol{\phi}}_1$ is a bi-dimensional vector formed by the first two vector components of $\hat{\boldsymbol{\phi}}_1$.

Note that evaluating (24) has complexity on the order of $O(M^2)$ [30]. On the other hand, the solution of (28) leverages operations on small-size matrices and vectors, with negligible computation complexity, especially in the presence of a large number of anchors.

3.4.2. Plane Intersection (PI) [27]

The rationale behind this approach is that any pair of TDoA measurements, coming from a group of three anchors, leads to an equation which

identifies the major axes of a conic, whose focus should lie in the target. Therefore, having at least three of these equations allows to determine the target position by solving the corresponding system of equations.[8]

The system of equations to be solved is

$$\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{b} \tag{30}$$

where $\boldsymbol{A}$ is a $(M-2)\times 2$ matrix and $\boldsymbol{b}$ is a length-$(M-2)$ vector defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} (x_1-x_2)\hat{\Delta}_3 - (x_1-x_3)\hat{\Delta}_2 & (y_1-y_2)\hat{\Delta}_3 - (y_1-y_3)\hat{\Delta}_2 \\ \vdots & \vdots \\ (x_1-x_2)\hat{\Delta}_M - (x_1-x_M)\hat{\Delta}_2 & (y_1-y_2)\hat{\Delta}_M - (y_1-y_M)\hat{\Delta}_2 \end{bmatrix} \tag{31}$$

$$\boldsymbol{b} = \begin{bmatrix} -\hat{\Delta}_2\,\hat{\Delta}_3(\hat{\Delta}_3-\hat{\Delta}_2) + (k_1^2-k_2^2)\hat{\Delta}_3 - (k_1^2-k_3^2)\hat{\Delta}_2 \\ \vdots \\ -\hat{\Delta}_2\,\hat{\Delta}_M(\hat{\Delta}_M-\hat{\Delta}_2) + (k_1^2-k_2^2)\hat{\Delta}_M - (k_1^2-k_M^2)\hat{\Delta}_2 \end{bmatrix}. \tag{32}$$

Hence, the LS solution can be computed as

$$\hat{\boldsymbol{u}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\left(\boldsymbol{A}^T\boldsymbol{b}\right) \tag{33}$$

with complexity on the order of $O(M^2)$ [30] (similar arguments to those for TSML can be applied).

3.4.3. Particle Swarm Optimization (PSO)

The considered geometric-based algorithms (TSML and PI) analytically solve the systems of equations described above. However, it may happen that the matrices involved in these systems of equations become ill-conditioned,

[8]Note that, in this case, a second reference anchor is needed to determine the system of equations. To this end, for notation simplicity, we will assume that $\boldsymbol{s}_2$ is the second anchor closest to the target.

thus leading to a very inaccurate target position estimate [31, 32]. The rationale behind the use of the PSO algorithm is to re-interpret the above systems of equations as minimization problems, avoiding numerical problems in finding their solutions. More precisely, the solution of the general localization problem (4) can be written as

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{p}} g(\boldsymbol{\rho}) \qquad \boldsymbol{\rho} \in \mathbb{R}^2 \tag{34}$$

where $g(\cdot)$ is the so-called fitness function and depends on the starting system of equations.

The fitness function should take into account the error, in the system of equations, when a wrong position is estimated: in particular, the larger the error, the higher the $g(\cdot)$. Starting from (21) for the TSML algorithm, the following fitness function can be considered:

$$g(\boldsymbol{\rho}) = \left\| \hat{\boldsymbol{\Delta}}^{\,2} - (\boldsymbol{k}^2 - k_1^2) + 2\,\hat{\boldsymbol{\Delta}}\,\|\boldsymbol{\rho} - \boldsymbol{s}_1\| + 2\,(\boldsymbol{S} - \boldsymbol{s}_1)^T\,\boldsymbol{p} \right\| \tag{35}$$

where $\hat{\boldsymbol{\Delta}}^{\,2}$ stands for the element-wise square of the vector $\hat{\boldsymbol{\Delta}}$. We refer to the PSO solution of (35) as PSO-TSML. Note that (35) provides a solution which is "biased" towards the first anchor $\boldsymbol{s}_1$, especially for large values $\hat{\boldsymbol{\Delta}}$, due to the presence of the term $2\,\hat{\boldsymbol{\Delta}}\,\|\boldsymbol{\rho} - \boldsymbol{s}_1\|$.

If, instead of the system given by (21), the system given by (30) is considered for the PI algorithm, the following fitness function can be used:

$$g(\boldsymbol{\rho}) = \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{\rho}\|\,. \tag{36}$$

We refer to the PSO solution of (36) as PSO-PI.

At this point, we resort to the standard implementation of the PSO (see, e.g., [33]) to solve the minimization problem (34) with fitness function either

equal to (35) or (36). According to the PSO algorithm, the set of potential solutions of each optimization problem, i.e., of the system of equations associated with the chosen localization algorithm, can be modeled as a swarm of particles. We denote the set of particles as $\mathcal{P}$ and its size as $|\mathcal{P}|$. The positions of the particles are randomly initialized in the region of interest and the key idea is to iteratively "guide" them towards the optimal solution by exploring the interactions between them.

At iteration[9] n ($n = 0, 1, \ldots, n_{\text{it}}$, where n_{it} is the number of iterations), the position and velocity of the i-th particle are represented by the two-dimensional vectors $\boldsymbol{\pi}_i[n]$ and $\boldsymbol{v}_i[n]$, respectively. The PSO algorithm assumes that each particle knows, at each iteration, its own best position (as will be discussed in the following) as well as the global best position among all the particles and the corresponding values of the fitness function.

The update rule for particles' velocities is given by

$$\begin{aligned} \boldsymbol{v}_i[n+1] &= \omega[n]\,\boldsymbol{v}_i[n] + c_1\,\chi_1[n]\,\{\boldsymbol{p}_i[n] - \boldsymbol{\pi}_i[n]\} \\ &\quad + c_2\,\chi_2[n]\,\{\boldsymbol{p}[n] - \boldsymbol{\pi}_i[n]\} \end{aligned} \tag{37}$$

where ω is the PSO inertial factor, c_1 and c_2 ($c_1, c_2 \in \mathbb{R}$, $c_1, c_2 \geq 0$) are the so-called *cognition* and *social* parameters, and $\chi_1[n]$ and $\chi_2[n]$ are independent random variables uniformly distributed in $[0, 1]$. Finally, $\boldsymbol{p}_i[n]$ and $\boldsymbol{p}[n]$ represent, respectively, the position of the i-th particle with the best fitness

[9]The use of an inherently iterative algorithm leads to a delay in the estimation procedure, which may be critical (depending on the available computation hardware) for real-time applications or dynamic scenarios with mobile targets. The investigation of refined techniques to reduce the number of iterations in conjunction with target tracking will be subject of future investigation.

function (over the n iterations) and the position of the particle with the best (among all particles) fitness function up to iteration n, i.e.,

$$\begin{aligned} \boldsymbol{p}_i[n] &= \underset{\boldsymbol{\rho}\in\{\boldsymbol{\pi}_i[j]\}_{j=0}^{n}}{\arg\min}\ g(\boldsymbol{\rho}) \\ \boldsymbol{p}[n] &= \underset{\boldsymbol{\rho}\in\{\boldsymbol{p}_i[n]\}_{i=1}^{|\mathcal{P}|}}{\arg\min}\ g(\boldsymbol{\rho}). \end{aligned}$$

Using (37), the update rule for the particles' positions is

$$\boldsymbol{\pi}_i[n+1] = \boldsymbol{\pi}_i[n] + \boldsymbol{v}_i[n+1].$$

In other words, the idea of the PSO algorithm is to check the system of equations in correspondence to some test positions, find the position with best fitness, and try to iteratively converge to the best position, by also exploring other positions in the surrounding space. The final solution is given by

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{\rho}\in\{\boldsymbol{p}_i[n_{\text{it}}]\}_{i=1}^{|\mathcal{P}|}}{\arg\min}\ g(\boldsymbol{\rho}).$$

As a final remark, one can note that the inertial factor $\omega[n]$ is representative of the ability of the particles to explore new areas in the surrounding space. However, taking into account the results in [31], in the following we will consider $\omega[n] \simeq 0\ \forall n$ in (37).

Since at each iteration the fitness function has to be evaluated for all particles, the complexity can be proved to be $O(|\mathcal{P}| \cdot n_{\text{it}} \cdot M)$ [30]. Since $|\mathcal{P}|, n_{\text{it}} \gg M$, the complexity of this algorithm may be unfeasible in most realistic applications that pose real-time requirements (e.g., those involved in IoT scenarios).

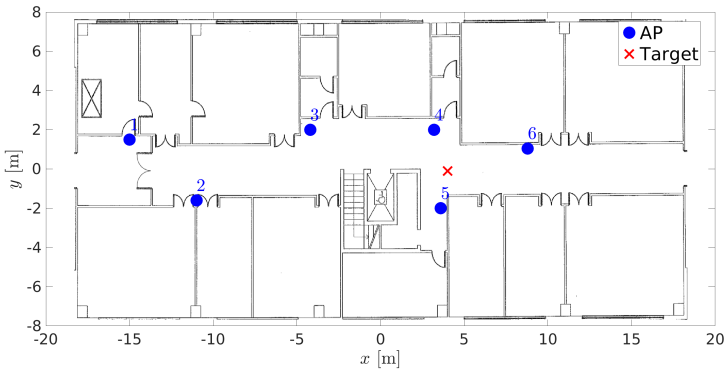


Figure 5: Indoor WiFi localization scenario.

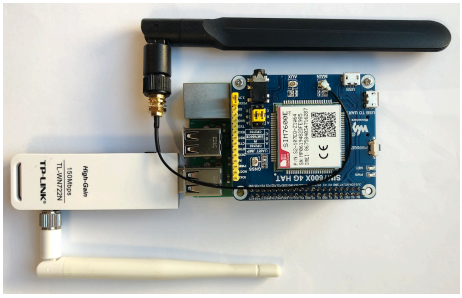


Figure 6: IoT target node used for the experimental performance analysis in indoor and outdoor localization scenarios.

4. Experimental Performance Analysis

We now present an experimental performance analysis in IEEE 802.11 (indoor) and LTE (outdoor) wireless scenarios. The indoor scenario is shown in Figure 5 and corresponds to the WiFi network deployed at the ground floor of the Building n. 2 of the Department of Engineering and Architecture of the University of Parma, Italy. The six anchors (denoted by blue circles) are IEEE 802.11 Cisco AIR-CAP3702I-E-K9 APs transmitting over 3 disjoint IEEE 802.11 20 MHz bandwidth channels (channels 1, 6, and 11). On the other hand, the target (denoted by a red cross and placed in an

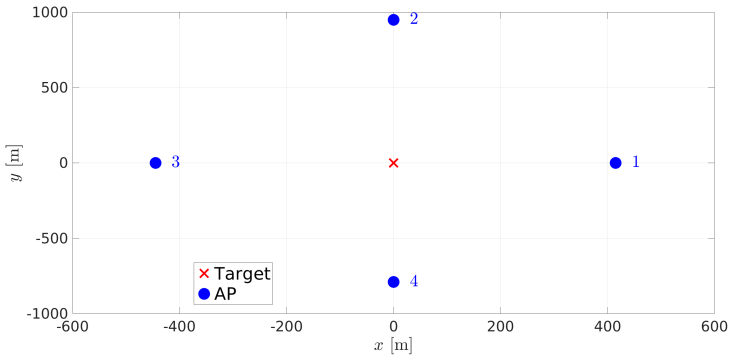


Figure 7: Outdoor LTE Localization Scenario. AP 1 and 2 are in LOS, AP 3 and 4 are in NLOS.

illustrative position) is shown in Figure 6 and corresponds to a Raspberry Pi 3 Model B+ (RPi) Single Board Computer (SBC), equipped with (i) an external IEEE 802.11 Linksys TL-WN722N USB dongle, with a 1 dB granularity RSSI measurement capability, and (ii) an *on-board*-plugged Waveshare SIM7600E-H 4G HAT expansion board, with a 1 dB granularity RSSI measurement capability. In the RPi, Wireshark is running to extract RSSI measurements from the beacon packets received by the APs. In particular, we consider a scenario with six APs, four of which are in LOS (those denoted as 3-4-5-6) and two are in NLOS (those denoted as 1-2) with respect to the target—LOS is associated with direct visibility, whereas NLOS is associated with the presence of at least one large object between target and anchor (e.g., wall, thick door, closet). Measurements are taken under different conditions (e.g., at various hours with different WiFi traffic loads, people passing by, etc.).

The outdoor LTE scenario is shown in Figure 7. The anchors are evolved Node Bs (eNBs) transmitting with an uplink frequency equal to 847 MHz

and a downlink frequency equal to 806 MHz, over a bandwidth of 10 MHz. The target corresponds to the aforementioned IoT node in Figure 6 running a proper script which collects RSSI data through the LTE hat's internal Application Programming Interfaces (APIs). Since the available APIs do not allow for simultaneous data collection from several eNBs, we collect single link measurements (in either LOS or NLOS conditions) and then build a "virtual" simulation scenario, as shown in Figure 7. The virtual simulation scenario is obtained by placing "virtual" eNBs at distances, from the target, equal to the distances from the real eNBs at which real experimental data were collected. Moreover, for each virtual eNB, LOS or NLOS channel status is selected on the basis of the corresponding experimental data: LOS is associated with direct visibility of anchors and target, whereas a NLOS status is associated with the presence of either the target inside a building or with the presence of at least one building between target and anchor. As for the indoor scenario, in this case as well measurements are taken under different environmental conditions for fixed target-eNBs' distances and LOS/NLOS conditions.

In both cases, data are collected at the target with a RSSI sampling interval of 100 ms. The number T of collected RSSI blocks from all the anchors is 842 for WiFi and 757 for LTE, each block containing $N = 30$ consecutive RSSI samples, corresponding to an observation time of 3 seconds. As preliminarily shown in [15], the accuracy of NLOS identification improves if N increases. Consequently, the performance of the mitigation can improve as well, since the NLOS links can be identified more accurately. However, increasing N leads to a longer delay in the acquisition process: this can be critical in dynamic scenarios with a mobile target. In this case, in fact, if

the acquisition delay is too long, the LOS/NLOS classification may refer to an outdated target position and may make the entire localization process inaccurate. Our results (with various values of N), not presented here for lack of space, show that $N = 30$ is a good trade-off between position estimation accuracy and delay. The number L of blocks of RSSI measurements associated with NLOS links is 278 for the WiFi scenario and 405 for the LTE scenario. Therefore, the number $T - L$ of blocks associated with LOS links is 564 for the WiFi case and 352 with LTE.

Channel classification is performed for each of the T blocks of RSSI measurements—channel status is considered constant over a block. When the reference NN classifier is considered, the input layer is the set of features $\tilde{\mathcal{D}}$, while the output corresponds to the estimated channel condition $\hat{\ell}$. Internally, the hidden layer is a 5-neuron fully connected layer using the scaled conjugate gradient training function. The size of the training, validation, and testing dataset are 80%, 10%, and 10% of T, respectively [15]. These values are chosen as a reasonable trade-off between training duration and NN performance. Moreover, cross-validation has been performed by considering different subsets' training, validation, and testing and, then, computing the average performance.t

Regarding channel mitigation, the regression weight vector $\boldsymbol{\gamma}$ in (20) is computed using all the L blocks of the NLOS measurements to generate the matrix $\boldsymbol{F}$ in (19). Then, this vector is used to compute the distance correction coefficient for the specific link and block according to (18).

Finally, a localization act is performed in correspondence to all time instants at which the target has been able to acquire a block of N RSSI mea-

surements from all the M anchors. The total number of performed localization acts, denoted as $K^{'}$, is around $100 \div 150$. The number $K^{'}$ is variable as it depends on the application (or not) of an outlier removal strategy. In particular, in this work we *a-priori* assume that the position estimate should lie inside the polytope identified by the anchors, i.e., the anchors should be at the boundary of the monitored area. Therefore, the considered outlier removal strategy eliminates position estimates that are significantly outside such a polytope.[10] In particular, we discard decisions for which $|\hat{x}_j| > 2 \max_i |\boldsymbol{s}_i|$ or $|\hat{y}_j| > 2 \max_i |\boldsymbol{s}_i|$, where $\hat{x}_j$ and $\hat{y}_j$ are the coordinates of the position estimate $\hat{\boldsymbol{u}}_j$ at the j-th localization act. For example, this could be appropriate for in-region user presence verification.

4.1. Indoor WiFi Scenario

We first analyze the performance of the NLOS classifier in terms of its Receiver Operating Characteristic (ROC) curve, defined as the probability of correct LOS classification, i.e., $P_{\mathrm{D}} = P(\hat{\ell} = 1 | \ell = 1)$, as a function of the probability of incorrect NLOS classification, i.e., $P_{\mathrm{F}} = P(\hat{\ell} = 1 | \ell = 0)$. The ROC curve for the indoor WiFi scenario is shown in Figure 8, considering various classifiers. For each considered classifier, the ROC curve is composed of points associated with different decision strategies, namely the values of single thresholds for the single-feature classifiers and $\tilde{\ell}_{\mathrm{th}}$ for the weighed classifier. The markers on the curves (one per curve) show the operational points in correspondence to which the final classification accuracy is approximately

[10] Improved outlier removal strategies can be considered, but this goes beyond the scope of this paper.

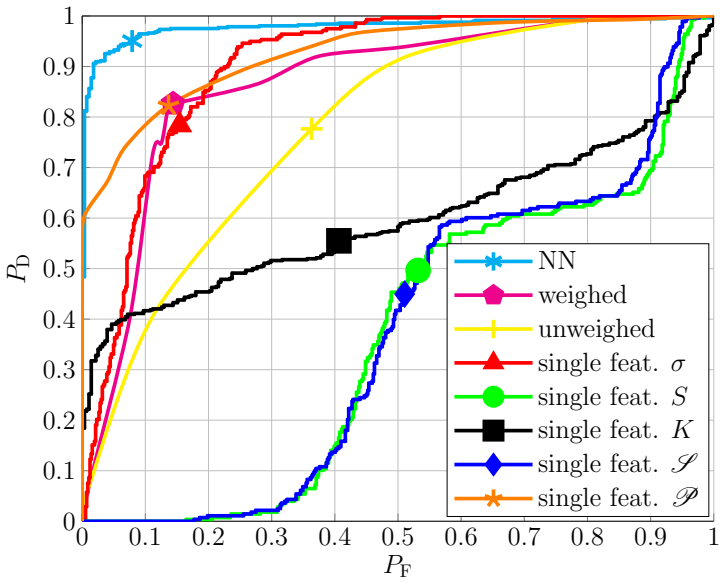


Figure 8: ROC curves of the considered classifiers in the indoor WiFi scenario. For each curve, the marker shows the working point for final classification accuracy of approximately 84%.

Table 1: Parameters for NLOS classification and mitigation in the indoor WiFi scenario.

feature	threshold	classification weights	mitigation weights ($\gamma_0 = -1.39$)
σ	$\sigma^* = 2$	$\alpha_1 = 0.2$	$\gamma_1 = 0.56$
S	$S^* = 0.6$	$\alpha_2 = 0.1$	$\gamma_2 = -1.59$
K	$\mathscr{S}^* = 5$	$\alpha_3 = 0.1$	$\gamma_3 = 0.31$
$\mathscr{S}$	$\mathscr{S}^* = 5$	$\alpha_4 = 0.1$	$\gamma_4 = 0.04$
$\mathscr{P}$	$\mathscr{P}^* = 0.7$	$\alpha_5 = 0.5$	$\gamma_5 = 3.40$

84%.[11] The values of the thresholds and weights of the various classifiers needed to achieve a final classification accuracy of approximately 84% are shown in Table 1. In this case, $\ell_{\text{th}} = 0.5$.

As can be seen from the results in Figure 8, the NN-based classifier provides the best solution, yet paying a higher price in terms of computational complexity and training needs. Moreover, for a final classification accuracy of 84%, weighing all the five considered features allows to achieve the best performance among the considered classifiers. However, the single-feature classifiers based on PP or σ achieve very good results (for instance, the performance with PP at the 84% accuracy overlaps with the performance with weighing) with a very simple implementation. This behavior makes single-feature classifiers attractive for energy conservation purposes (especially for

[11] This value is chosen as a reasonable performance trade-off.

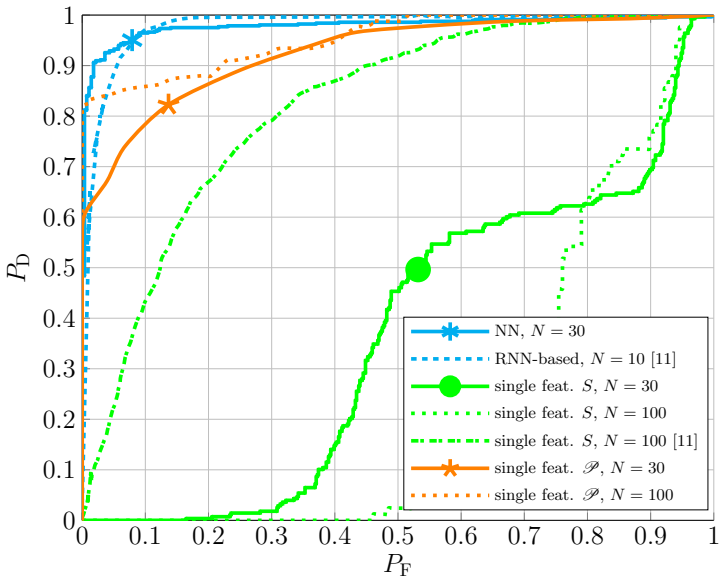


Figure 9: ROC curves of some of the proposed classifiers in the indoor WiFi scenario and various values of N. A comparison with the classifiers in [16] is provided.

mobile and constrained devices, such as for IoT-oriented scenarios). On the other hand, other single-feature classifiers (based on S, K and $\mathscr{S}$) do not provide acceptable performance.

In Figure 9, we compare our classification results in the indoor WiFi scenario (considering two values of N, namely 30 and 100) with those in [16], where the following approaches have been considered: (i) a reference case with Recurrent Neural Network (RNN) and $N = 10$ and (ii) a single statistical feature (namely the skewness) classifier with $N = 100$. One can observe that the NN (ours) and RNN ([16]) benchmarks have similar performance. Moreover, our single-feature classifier (based on the PP) with $N = 30$ outperforms single-feature classifier (based on skewness) of [16] with $N = 10$. Note that our accuracy, precision, and sensitivity results are in agreement

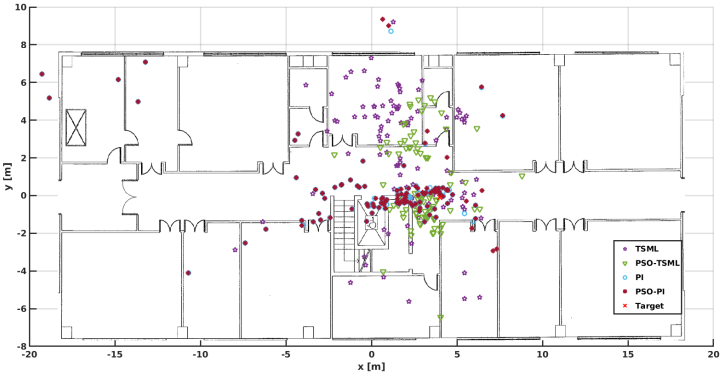


Figure 10: Estimated positions in indoor WiFi scenario with outlier removal.

with those in [22, Table 1].

After classification, NLOS mitigation is carried out. In Table 1, the weights for the mitigation procedure, corresponding to each feature, are shown.

After mitigation of the 2 NLOS links associated to anchors 1 and 2 (see Figure 5), localization is carried out. The estimated positions are shown in Figure 10. As a concise performance indicator of the localization accuracy, we evaluate the Root Mean Square Error (RMSE, dimension: [m]), defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{K'} \sum_{j=1}^{N} |\boldsymbol{u} - \hat{\boldsymbol{u}}_j|^2}.$$

The RMSE results are shown in Figure 11 for the indoor WiFi scenario. The case without and with NLOS mitigation are compared. In the latter case, we show the results for both the cases without and with outlier removal. One can observe that our (pre-localization) NLOS mitigation strategy significantly reduces the RMSE, especially if carried out with outlier removal. The only exception is the PSO-TMSL algorithm, which performs well also in

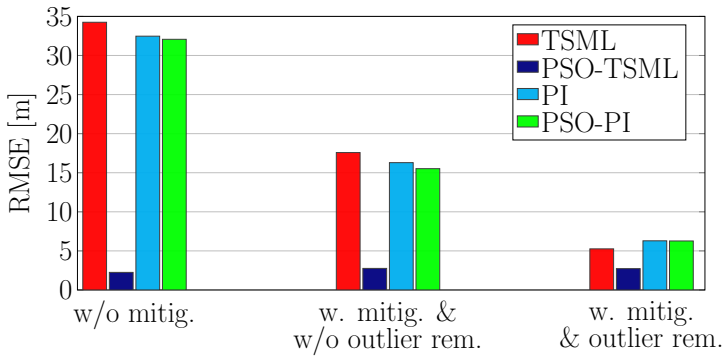


Figure 11: RMSE for the indoor WiFi scenario.

the absence of correction. However, as already previously observed, the complexity of a PSO-based solution is typically high and, therefore, the adoption of such a solution is unfeasible in IoT scenarios. Moreover, PSO requires an iterative process, which may also hinder its applicability from a delay point of view. However, PSO-TSML represents a relevant performance benchmark.

4.2. Outdoor LTE Scenario

We now analyze the performance in the outdoor LTE scenario. The ROC curves for the considered classifiers are shown in Figure 12. Considerations similar to those carried out for Figure 8 are still valid, except for the fact that, in general, link classification and, therefore, NLOS mitigation is less accurate. In Figure 12, the operational points indicated with a symbol over the ROC curves correspond to a final localization accuracy of 75% (the highest with LTE). As can be seen, the weighed classifiers (both using all features) allow to achieve the best performance (for a final accuracy of 75%), whereas the performance with the single-feature classifiers degrades. Therefore, the weighed classifier can be considered as "universal" in the sense that it al-

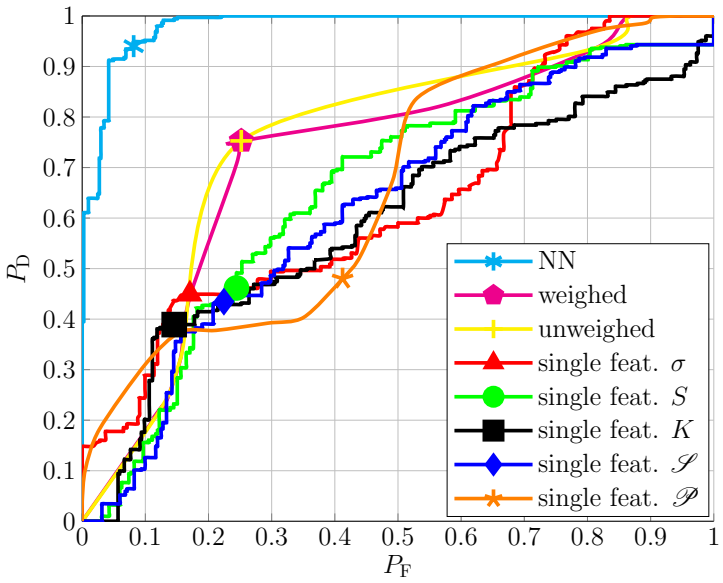


Figure 12: ROC curves of the considered classifiers in the outdoor LTE scenario. For each curve, the marker shows the working point for final classification accuracy of approximately 75%.

Table 2: Parameters for NLOS classification and mitigation in the outdoor LTE scenario.

feature	threshold	classification weights	mitigation weights ($\gamma_0 = -0.62$)
σ	$\sigma^* = 1.7$	$\alpha_1 = 0.2$	$\gamma_1 = 5.58$
S	$S^* = 0$	$\alpha_2 = 0.2$	$\gamma_2 = -1.83$
K	$\mathscr{S}^* = 3$	$\alpha_3 = 0.3$	$\gamma_3 = 6.51$
$\mathscr{S}$	$\mathscr{S}^* = -0.1$	$\alpha_4 = 0.2$	$\gamma_4 = -7.88$
$\mathscr{P}$	$\mathscr{P}^* = 0.7$	$\alpha_5 = 0.1$	$\gamma_5 = 0.71$

lows to achieve almost the best performance in both (indoor and outdoor) considered scenarios. The thresholds and weights of the NLOS classifier, to achieve a final classification accuracy of approximately 75%, are shown in Table 2, together with the corresponding mitigation weights. In this case as well, $\ell_{\text{th}} = 0.5$.

The estimated positions and the RMSE for the outdoor LTE scenario are shown in Figure 13 and Figure 14, respectively. From the results in Figure 14, it can be observed that, even if our mitigation strategy drastically reduces the RMSE, the localization error seems still significant (on the order of 400 m in the best case). This is probably due to the fact that using only four anchors (eNBs) is not sufficient in LTE scenarios. In this case as well, the only exception is the PSO-TMSL algorithm, which works very well also in the absence of NLOS link mitigation. While in IoT scenarios the complexity

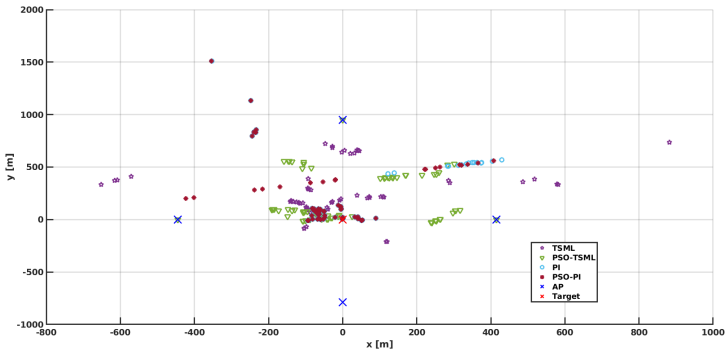


Figure 13: Estimated positions in the outdoor LTE scenario with outlier removal.

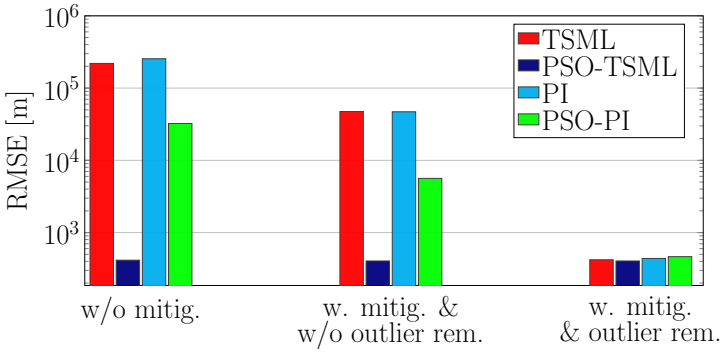


Figure 14: RMSE for the outdoor LTE scenario.

of PSO-TSML prevents its use, this may be attractive in cellular (4G/5G) scenarios.

4.3. Indoor/Outdoor Scenario Comparison

We finally compare directly the performance of the proposed localization method, with NLOS pre-mitigation, in the considered outdoor (LTE) and indoor (WiFi) scenarios. In Figure 15, the ROC curves for the outdoor LTE and indoor WiFi scenarios are directly compared, considering both the reference NN and the best (weighed) classifier. One can observe that the

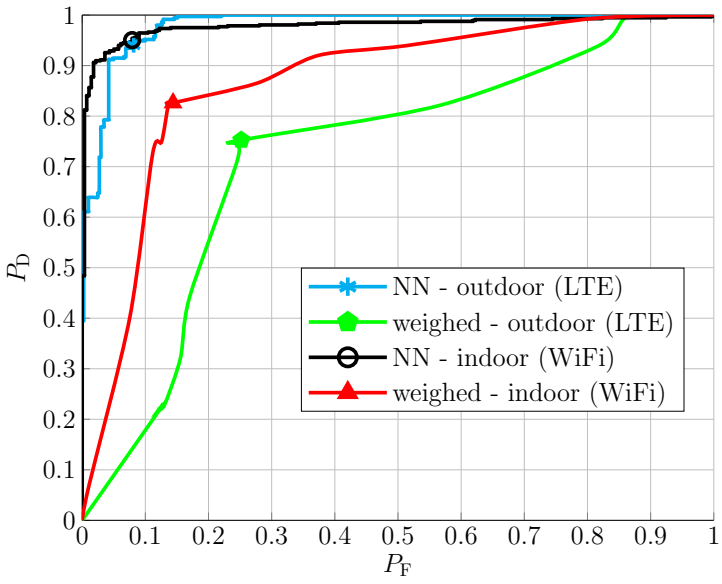


Figure 15: Comparison of the ROC curves for the outdoor LTE and indoor WiFi scenarios, considering both the reference NN and the best (weighed) classifier.

reference NN classifier has approximately the same performance classification in indoor and outdoor scenarios. On the other hand, the weighed classifier has better performance in the indoor case. The shapes of the ROC curves with the weighed classifier are the same in indoor and outdoor scenarios.

We now set to compare the localization accuracy in outdoor and indoor scenarios. In order to make a fair comparison, we "normalize" the RMSE with respect to the average distance between target and anchors, which is 650 m in the outdoor scenario and approximately 8.62 m in the indoor scenario. Therefore, the normalized RMSE is a relative measure of the localization estimation error with respect to the considered topology. In Figure 16, we compare the normalized RMSEs in outdoor LTE (from Figure 14) and indoor WiFi (from Figure 11) scenarios: in all cases, NLOS pre-mitigation,

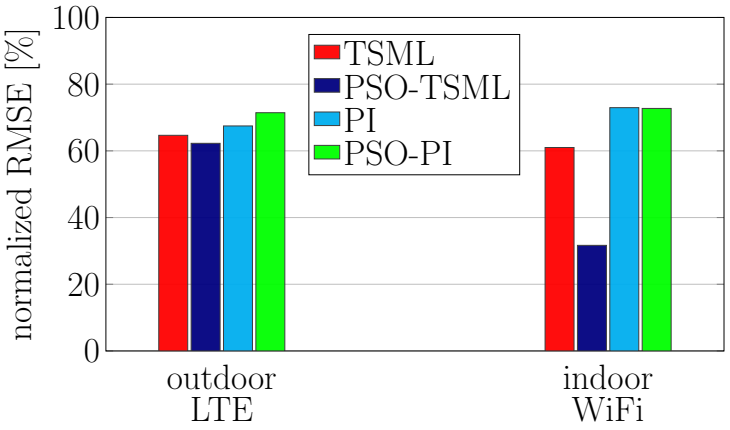


Figure 16: Comparison of the normalized RMSE for the outdoor LTE and indoor WiFi scenarios.

together with outlier removal, is considered. It is worth noting that, except for the PSO-TSML, the relative RMSE is approximately the same for all considered algorithms in both scenarios. This is a desirable feature of the proposed strategy, since it means that its performance is not affected by the specific scenario. Intuitively, this can be justified by the fact that communication bandwidths are similar in outdoor and indoor scenarios and it is well-known that the localization error is approximately inversely proportional to such a quantity [34]. A thorough analytical demonstration of this behavior is the subject of on-going research.

5. Concluding Remarks

In this paper, we have proposed a general approach to RSSI-based localization. The availability of RSSI from COTS devices makes this approach attractive for IoT scenarios. The key idea is to leverage channel status identification (LOS/NLOS) to perform NLOS mitigation before localization. In

particular, classification and mitigation are based on the computation of significant statistical features over an observation window of N consecutive RSSI samples. Our results show that good channel status classification accuracy can be achieved by simply using a threshold detector based on a single statistical feature, namely the PP. Mitigation is then performed by determining heuristic correction coefficients (depending on the links' statistical features) for the distance estimates associated with NLOS links: simply put, a NLOS link is transformed into an equivalent LOS one. A few TDoA-based localization algorithms have then been considered and experimental performance assessment has been carried out in indoor (WiFi) and outdoor (LTE) scenarios. Our results show that (low-complexity) localization algorithms (namely, TSML and PI) significantly benefit in terms of position error reduction with respect to an approach which does not exploit any NLOS mitigation. In general, TSML-PSO guarantees the best performance in all considered scenarios, regardless of NLOS link mitigation: however, its applicability is not feasible in IoT scenarios, but may be attractive in 4G/5G systems. Finally, the approach seems to be universal in the sense that the relative localization error, with respect to the average target-anchors' distance, is approximately the same in outdoor or indoor scenarios.

References

[1] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, G. Seco-Granados, Survey of cellular mobile radio localization methods: From 1G to 5G, IEEE Communications Surveys Tutorials 20 (2) (2018) 1124–1148, DOI: 10.1109/COMST.2017.2785181.

[2] Y. Li, Y. Zhuang, X. Hu, Z. Gao, J. Hu, L. Chen, Z. He, L. Pei, K. Chen, M. Wang, X. Niu, R. Chen, J. Thompson, F. Ghannouchi, N. El-Sheimy, Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation, IEEE Internet of Things J.DOI: 10.1109/JIOT.2020.3019199.

[3] F. Zafari, A. Gkelias, K. Leung, A survey of indoor localization systems and technologies, IEEE Commun. Surveys Tutorials 21 (3) (2019) 2568–2599, DOI: 10.1109/COMST.2019.2911558.

[4] Y. Zhao, Z. Zhang, T. Feng, W. Wong, H. K. Garg, GraphIPS: Calibration-free and map-free indoor positioning using smartphone crowdsourced data, IEEE Internet of Things J.DOI: 10.1109/JIOT.2020.3004703.

[5] S. Sadowski, P. Spachos, K. N. Plataniotis, Memoryless techniques and wireless technologies for indoor localization with the Internet of Things, IEEE Internet of Things J.DOI: 10.1109/JIOT.2020.2992651.

[6] L. Chen, I. Ahriz, D. Le Ruyet, AoA-aware probabilistic indoor location fingerprinting using channel state information, IEEE Internet of Things J.DOI: 10.1109/JIOT.2020.2990314.

[7] Y. Li, X. Hu, Y. Zhuang, Z. Gao, P. Zhang, N. El-Sheimy, Deep reinforcement learning (DRL): Another perspective for unsupervised wireless localization, IEEE Internet of Things J. 7 (7) (2020) 6279–6287, DOI: 10.1109/JIOT.2019.2957778.

[8] J. Talvitie, M. Renfors, M. Valkama, E. S. Lohan, Method and analysis of spectrally compressed radio images for mobile-centric indoor localization, IEEE Trans. Mobile Comput. 17 (4) (2018) 845–858, DOI:10.1109/TMC.2017.2741487.

[9] Z. Kasmi, N. Guerchali, A. Norrdine, J. H. Schiller, Algorithms and position optimization for a decentralized localization platform based on resource-constrained devices, IEEE Trans. Mobile Comput. 18 (8) (2019) 1731–1744, DOI:10.1109/TMC.2018.2868930.

[10] S. Kumar, S. K. Das, Target detection and localization methods using compartmental model for Internet of Things, IEEE Trans. Mobile Comput. 19 (9) (2020) 2234–2249, DOI:10.1109/TMC.2019.2921537.

[11] S. Shao, A. Khreishah, I. Khalil, Enabling real-time indoor tracking of IoT devices through visible light retroreflection, IEEE Trans. Mobile Comput. 19 (4) (2020) 836–851, DOI:10.1109/TMC.2019.2901665.

[12] F. Shahzad, T. R. Sheltami, E. M. Shakshuki, DV-maxHop: A fast and accurate range-free localization algorithm for anisotropic wireless networks, IEEE Trans. Mobile Comput. 16 (9) (2017) 2494–2505, DOI:10.1109/TMC.2016.2632720.

[13] Y. Zhao, X. Li, Y. Ji, C. Xu, Wireless power-driven positioning system: Fundamental analysis and resource allocation, IEEE Internet of Things J. 6 (6) (2019) 10421–10430, DOI: 10.1109/JIOT.2019.2939215.

[14] Z. Shi, Y. Wang, User positioning by exploring MIMO measurements with particle swarm optimization, in: Proc. IEEE Intern.

Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC), Montreal, Canada, 2017, pp. 1–5, DOI:10.1109/PIMRC.2017.8292644. doi:10.1109/PIMRC.2017.8292644.

[15] F. Carpi, L. Davoli, M. Martalò, A. Cilfone, Y. Yu, Y. Wang, G. Ferrari, RSSI-based methods for LOS/NLOS channel identification in indoor scenarios, in: Proc. Int. Symp. Wireless Communication Systems (ISWCS), Oulu, Finland, 2019, pp. 171–175, DOI: 10.1109/ISWCS.2019.8877315.

[16] J. Choi, W. Lee, J. Lee, J. Lee, S. Kim, Deep learning based NLOS identification with commodity WLAN devices, IEEE Trans. Veh. Technol. 67 (4) (2018) 3295–3303, DOI: 10.1109/TVT.2017.2780121.

[17] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, Y. Bo, UWB NLOS/LOS classification using deep learning method, IEEE Commun. Lett. 24 (10) (2020) 2226–2230, DOI:10.1109/LCOMM.2020.2999904.

[18] S. Zhang, C. Yang, D. Jiang, X. Kui, S. Guo, A. Y. Zomaya, J. Wang, Nothing blocks me: Precise and real-time LOS/NLOS path recognition in RFID systems, IEEE Internet of Things J. 6 (3) (2019) 5814–5824, DOI: 10.1109/JIOT.2019.2907555.

[19] S. Maranò, W. M. Gifford, H. Wymeersch, M. Z. Win, NLOS identification and mitigation for localization based on UWB experimental data, IEEE J. Select. Areas Commun. 28 (7) (2010) 1026–1035, DOI: 10.1109/JSAC.2010.100907.

[20] Y. Chen, S. Huang, T. Wu, W. Tsai, C. Liou, S. Mao, UWB system for indoor positioning and tracking with arbitrary target orientation,

optimal anchor location, and adaptive NLOS mitigation, IEEE Trans. Veh. Technol.DOI: 10.1109/TVT.2020.2972578.

[21] X. Cai, X. Li, R. Yuan, Y. Hei, Identification and mitigation of NLOS based on channel state information for indoor WiFi localization, in: Proc. Int. Conf. Wireless Communications Signal Processing (WCSP), Nanjing, China, 2015, pp. 1–5, DOI: 10.1109/WCSP.2015.7341172.

[22] K. Bregar, M. Mohorčič, Improving indoor localization using convolutional neural networks on computationally restricted devices, IEEE Access 6 (2018) 17429–17441, DOI: 10.1109/ACCESS.2018.2817800.

[23] M. Katwe, P. Ghare, P. K. Sharma, A. Kothari, NLOS error mitigation in hybrid RSS-TOA based localization through semi-definite relaxation, IEEE Commun. Lett.DOI: 10.1109/LCOMM.2020.3020948.

[24] Y. Wang, K. Gu, Y. Wu, W. Dai, Y. Shen, NLOS effect mitigation via spatial geometry exploitation in cooperative localization, IEEE Trans. Wireless Commun. 19 (9) (2020) 6037–6049, DOI: 10.1109/TWC.2020.2999667.

[25] A. Goldsmith, Wireless Communications, Cambridge University Press, New York, NY, USA, 2005.

[26] Y. Chan, K. C. Ho, A simple and efficient estimator for hyperbolic location, IEEE Trans. Signal Processing 42 (8) (1994) 1905—1915, DOI: 10.1109/78.301830.

[27] R. O. Schmidt, A new approach to geometry of range difference location,

IEEE Trans. Aerosp. Electron. Syst. AES-8 (6) (1972) 821–835, DOI: 10.1109/TAES.1972.309614.

[28] K. Cengiz, Comprehensive analysis on least squares lateration for indoor positioning systems, IEEE Internet of Things J.DOI: 10.1109/JIOT.2020.3020888.

[29] Å. Björck, Linear Algebra and its Applications, Elsevier Science Publishers, 1987.

[30] H. Cui, Y. Liang, C. Zhou, N. Cao, Localization of large-scale wireless sensor networks using Niching particle swarm optimization and reliable anchor selection, Hindawi Wireless Commun. Mobile Comput.Article ID 2473875, DOI: 10.1155/2018/2473875.

[31] S. Monica, G. Ferrari, Swarm intelligent approaches to auto-localization of nodes in static UWB networks, Applied Soft Computing 25 (2014) 426–434, DOI: 10.1016/j.asoc.2014.07.025.

[32] S. Monica, G. Ferrari, Maximum likelihood localization: When does it fail?, ICT Express 2 (1) (2016) 10 – 13, DOI: 10.1016/j.icte.2016.02.004.

[33] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proc. Int. Conf. Neural Networks (ICNN), Vol. 4, Perth, WA, Australia, 1995, pp. 1942–1948, DOI: 10.1109/ICNN.1995.488968.

[34] S. Gezici, Zhi Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, Z. Sahinoglu, Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks, IEEE Signal Processing Mag. 22 (4) (2005) 70–84, DOI:10.1109/MSP.2005.1458289.

Fabrizio Carpi received his M.Sc. degree in Communication Engineering from University of Parma, Italy, in 2018. He is currently a Ph.D. candidate in Electrical Engineering at New York University Tandon School of Engineering. He is also a member of NYU WIRELESS Center conducting research on next-generation wireless networks. His research interests include wireless communications, source and channel coding, information theory, and machine learning.

Marco Martalò is an Associate Professor of Telecommunications at the University of Cagliari, Italy, which he joined in 2020 and where he is part of the Networks for Humans (Net4U) laboratory. From 2012 to 2017, he was an Assistant Professor with E-Campus University, Italy, and also a Research Associate with the University of Parma, Italy, until 2020. He has co-authored the book "Sensor Networks with IEEE 802.15.4 Systems: Distributed Processing, MAC, and Connectivity." His research interests are in the design of communication and signal processing algorithms for wireless systems and networks.

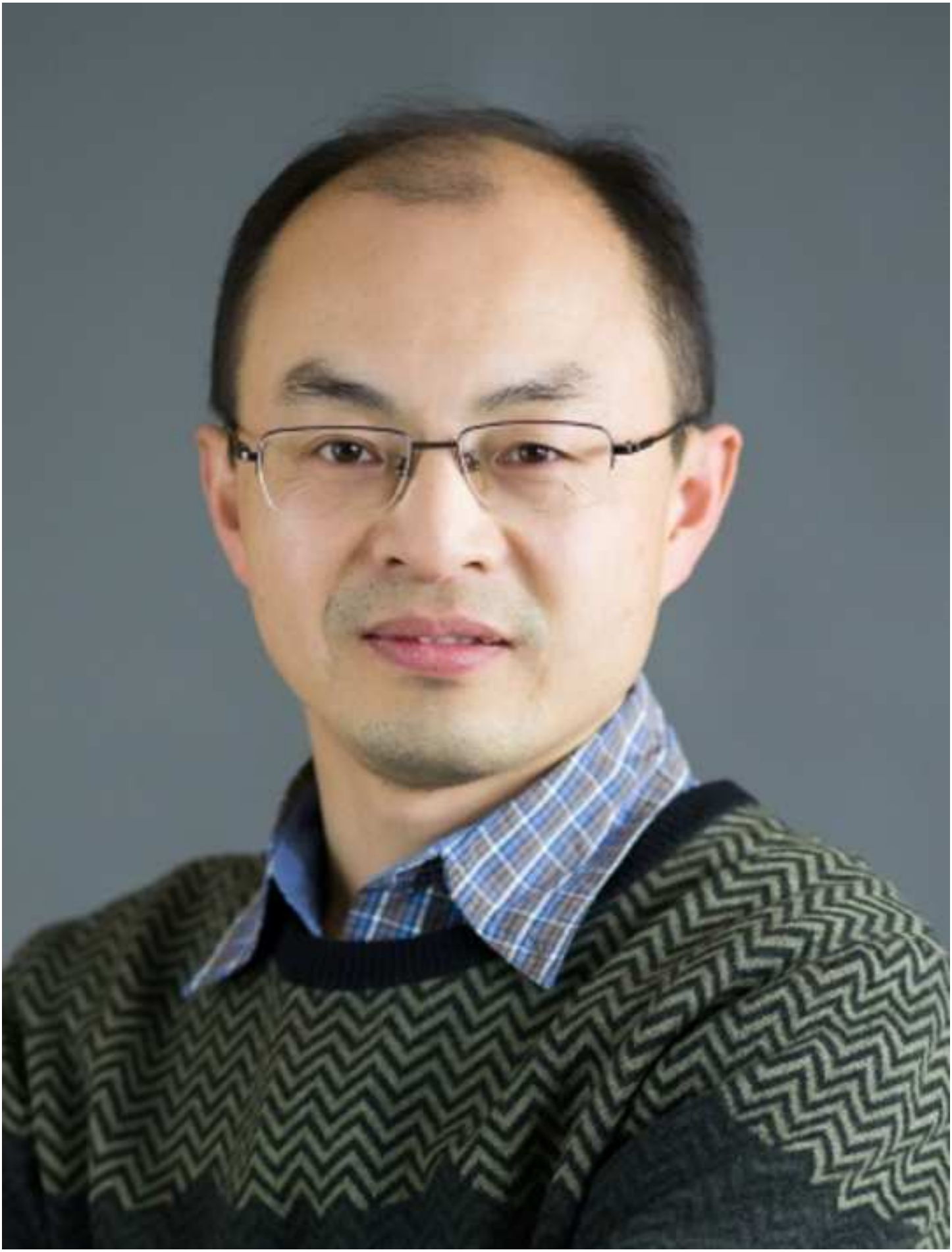
Luca Davoli is currently a Fixed-term Assistant Professor at the Department of Engineering and Architecture of the University of Parma. He received his Ph.D. in Information Technologies in 2017 and, since January 2014, he has been a member of the Internet of Things (IoT) Laboratory at the same university. He has (published or in press) over 35 papers, and has been TPC member of international conferences and served on the editorial boards and as Guest Editor of international journals. His main research interests are in the fields of Internet of Things, Software Defined Networking, Big Stream and Peer-to-Peer Networks.

Antonio Cilfone received his Master of Science in Communication Engineering and his Ph.D. in Information Technologies from the University of Parma, Parma, Italy, in 2016 and 2019, respectively. He has been member of the Internet of Things (IoT) Lab at the Department of Engineering and Architecture of the University of Parma from 2016 until 2020, working in heterogenous networking, signal processing and smart systems fields. He is currently working as R&D software engineer for Tesmec Automation S.r.l., Italy.

Yingjie YU received the B.S., M.S., and Ph.D. degrees in communication and information system from Northwestern Polytechnical University, Xi'an, China, in 2011, 2014, and 2018, respectively. She is currently a senior engineer in Huawei Technologies Co.,Ltd working on 5G positioning research.

Dr. Yi Wang is currently a principal engineer at Huawei Technologies Co., Ltd. in Shanghai. Since 2005 he joined Huawei Technologies Co., Ltd. he led a series of research projects on LTE/LTE-Advanced and 5G. Currently he is leading 5G positioning research in Huawei. Dr. Yi Wang owns granted 150+ patents and 90+ papers. Many patents have been realized in LTE/LTE-Advanced and 5G products and adopted in 3GPP and IEEE802.11 standards. Dr. Yi Wang is the board member of NYU Wireless Industrial Affiliates since 2014. He is the chair of China IMT-2020 (5G) mmWave Technology since 2013.

Gianluigi Ferrari (http://www.tlc.unipr.it/ferrari) is an Associate Professor of Telecommunications at the University of Parma, Italy, where he coordinates the Internet of Things (IoT) Lab (http://iotlab.unipr.it/) in the Department of Engineering and Architecture. His research interests revolve around three main areas: advanced communication and networking; signal processing; IoT and smart systems. He is an IEEE Senior Member. Since 2016, he is Co-founder and President of things2i ltd. (http://www.things2i.com/), a spin-off company of the University of Parma dedicated to IoT and smart systems. In 2021, he has been included in the world's top 2% of Scientists List by Stanford University.



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: