



# A Pre-processing Algorithm Utilizing a Paired CRLB for TDoA Based IoT Positioning

Ahmed Abdel Ghany, Bernard Uguen, Dominique Lemur

## ► To cite this version:

Ahmed Abdel Ghany, Bernard Uguen, Dominique Lemur. A Pre-processing Algorithm Utilizing a Paired CRLB for TDoA Based IoT Positioning. 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), May 2020, Antwerp, Belgium. pp.9128385, 10.1109/VTC2020-Spring48590.2020.9128385 . hal-03083166

**HAL Id: hal-03083166**

**<https://hal.science/hal-03083166>**

Submitted on 18 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Pre-processing Algorithm Utilizing a Paired CRLB for TDoA Based IoT Positioning

Ahmed Abdel Ghany, Bernard Uguen, Dominique Lemur  
*Institute of Electronics and Telecommunications of Rennes (IETR)*

*University of Rennes 1*

Rennes, France

{ahmed.abdelghany, bernard.uguen, dominique.lemur}@univ-rennes1.fr

**Abstract**—Time Difference of Arrival (TDoA) is currently viewed as an important technique for the positioning capabilities in the Internet of Things (IoT). However, in the case of practical measurement, not all the TDoA values between the gateways have the same impact on the localization accuracy. In this paper, a novel TDoA pre-processing methodology for dropping out the outlier TDoA values is presented, after instrumentalizing a paired Cramér-Rao lower bound (CRLB). Thus, the proposed approach is detecting the best TDoA values, which have the lowest paired CRLB values specifically, in the vicinity of the guessed node location, based on a robust thresholding method. A comparison is performed investigating the attainable accuracies for localizing based on this pre-processing algorithm, on a well-defined simulation environment. This simulator is based on a Poisson distribution approach for defining the gateways and node positions, as well as a noise model for emulating the timestamp imperfections. In the given results, the feasibility of the proposed technique is asserted by a drastic improvement over a wide range of the number of gateways as well as measurement noise variances. This manifests the robustness of the contributed method to the outlier TDoA values and its valuable rendering.

**Index Terms**—IoT, LoRa, Localization, TDoA, CRLB

## I. INTRODUCTION

Internet of Things (IoT) has been scaling up over the last few years as a new paradigm enabling to connect trillions of devices efficiently [1]. Beyond the traditional voice, video and data services where data throughput is the main purpose in the context of the IoT, the focus here is on low-cost deployments with large coverage areas. For providing this connectivity, Low Power Wide Area Networks (LPWAN) are considered the major technology, especially the LoRaWAN whose design compromises between low energy consumption and a large communication range of the underlying IoT-enabled devices. This exhibited in the typical LPWAN devices, which can cover distances of more than 10 km using a 10 mW transmit powers while maintaining extremely long battery lifetimes at the scale of up to several years [2].

Due to the very nature of the data collected from the terminals, positioning is critical for many LPWAN applications. These usages involve health data, monitoring of pets or livestock, wildlife or applications for smart cities using sensors such as temperature or air quality monitoring in urban environments [3]. This problem could be simply

solved by equipping each sensor with a Global Navigation Satellite Systems (GNSS) chip, for example, using GPS. This solution is tempting, however, adding a GPS tracker to a device will increase both cost and power consumption [4]. Subsequently, plenty of researches within the domain of GPS-Free localization in IoT, have investigated methods based on Received Signal Strength Indicator (RSSI), Angle of Arrival (AoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA) and their multiple integrations [5]. These measurements are utilized by the gateways to determine their relative position relations with the source for localization. However, using the RSSI, AoA and ToA techniques always require a good knowledge of the signal attenuation model, precise calibration or additional hardware [6], [7]. Therefore, these three kinds of techniques are not so practical.

TDoA is considered the most popular technique for localization as it does not require the transmitter to be synchronized with the receivers but only the gateways are required to have synchronized clocks [8]. This accurate time synchronization can be achieved, easily and cheaply, by the use of a GPS receiver at each gateway. Thus, a mobile node transmits data to the network. Each gateway within reach records the timestamp of the received packet. The timestamps of each gateway are then forwarded to the network server which in turn sends a request to the geolocation solver. After acquiring the differences between the timestamps of a transmission, it is considered a multilateration problem which involves solving a set of hyperbolic functions, and therefore three gateways are needed to locate the node on the intersecting hyperbolas. Based on that, the conventional geolocation for LoRa networks can be based on TDoA measurements, where hundreds to thousands of meters accuracy can be achieved. However, in a practical situation, not all pairs of gateways are equivalent concerning the amount of extra information they bring to the estimation problem. Using the entire combinations as the classical TDoA solving techniques may have an extremely deleterious effect on the final position accuracy. Thus, it could be advantageous to define some criteria to select the best combination of gateways or equivalently the best hyperbolas to resolve the estimation problem.

In this paper, to tackle this issue, a novel pre-processing method is proposed by utilizing a paired Cramér-Rao lower bound (CRLB) as an instrumental tool to detect the mostly perturbed TDoA values. This capability is due to the relative position of the node with respect to each pair of gateway locations which affect the paired CRLB values specifically. Accordingly, the perturbation probability in each TDoA value can vary largely. The accuracy and ease of use of this pre-processing algorithm are evaluated through simulations. It is shown that the proposed approach effectively reduces localization errors in a wide range of situations covering different densities of gateways and measurement noise variances.

The remainder of this document is organized as follows. Section II presents the system model and Section III provides sufficient detail to allow implementation of the proposed algorithm. Section IV provides guidance on the simulation model used to evaluate the performance of the algorithm on the localization accuracy. The results of the simulation are then presented and commented in Section V. Finally, Section VI is dedicated for conclusions.

## II. SYSTEM MODEL

A set of  $N$  gateways are considered with known positions

$$\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_n, \dots, \mathbf{g}_N] \quad (1)$$

with

$$\mathbf{g}_n = [x_n, y_n]^T, \quad (2)$$

while  $\mathbf{p} = [x, y]^T$  is any node position in the plane. Thus, the timestamp of the transmitted packet from  $\mathbf{p}$  is measured on gateway  $n$  as:

$$t_n = \frac{d_n}{c} + u_n = \tau_n + u_n, \quad (3)$$

where  $c$  is the celerity of light and  $u_n \sim \mathcal{N}(\delta, \sigma^2)$  is an additive Gaussian random variable which is accounting for the departure between the timestamp  $t_n$  and the node to gateway time of flight  $\tau_n$ . Notice that the standard deviation  $\sigma$  could be made dependent on the gateway index but this dependency is here omitted. Moreover, the time offset  $\delta$  is assumed to be equal for all gateways, and Line of sight (LoS) conditions are implicitly assumed while,  $d_n = \|\mathbf{g}_n - \mathbf{p}\|_2$  is the distance between any node position and the gateway  $n$  where,  $\|\cdot\|_2$  denotes the 2-norm. Each combination of 2 among  $N$  gateways leads to:

$$H = \binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2} \quad (4)$$

constraints, each being associated with one hyperbola. An hyperbola  $j$  involves the 2 gateways  $l(j)$  and  $r(j)$  while,  $\mathbf{l} = [l(1), l(2), \dots, l(H)]$  and  $\mathbf{r} = [r(1), r(2), \dots, r(H)]$ . Accordingly,

$$\mathbf{G}_\mathbf{r} = [\mathbf{g}_{r(1)}, \dots, \mathbf{g}_{r(H)}], \quad (5)$$

$$\mathbf{G}_\mathbf{l} = [\mathbf{g}_{l(1)}, \dots, \mathbf{g}_{l(H)}], \quad (6)$$

$$h_{l(i),r(i)} = d_{l(i)} - d_{r(i)}, \quad (7)$$

$$\tau_{l(i),r(i)} = t_{l(i)} - t_{r(i)}, \quad (8)$$

$$\mathbf{h} = [h_{l(1),r(1)}, \dots, h_{l(H),r(H)}] \quad (9)$$

and

$$\boldsymbol{\delta}_\tau = [\tau_{l(1),r(1)}, \dots, \tau_{l(H),r(H)}]. \quad (10)$$

In the classical TDoA techniques, the estimated node location  $\hat{\mathbf{p}} = [\hat{x}, \hat{y}]^T$  is obtained using the conventional algorithms such as the Least Squares (LS) or under Gaussianity assumption by Maximum Likelihood estimation (ML) [9]:

$$\hat{\mathbf{p}}_{ML} = \arg \min_{x,y} ((\mathbf{h} - c\boldsymbol{\delta}_\tau)^T \mathbf{R}^{-1} (\mathbf{h} - c\boldsymbol{\delta}_\tau)) \quad (11)$$

with the associated  $H \times H$  covariance matrix:

$$\mathbf{R} = \begin{bmatrix} 2\sigma^2 & \sigma^2|0 & \dots & \dots \\ \sigma^2|0 & 2\sigma^2 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \dots & 2\sigma^2 \end{bmatrix}, \quad (12)$$

where the notation  $\sigma^2|0$  means that each off-diagonal term of  $\mathbf{R}$  is either equal to  $\sigma^2$  or 0 provided that the two couples of gateways i.e., involved in that term, share respectively either one gateway or none.

## III. PAIRED CRLB PRE-PROCESSING

The Cramér-Rao lower bound (CRLB) is a lower bound on the covariance of any unbiased estimation algorithm based on the measurement [10]. This is calculated for the whole TDoA combinations from the inverse of the Fisher Information Matrix (FIM)  $\mathbf{J}$  as :

$$E[(\hat{\mathbf{p}} - \mathbf{p}_t)(\hat{\mathbf{p}} - \mathbf{p}_t)^T] \geq \mathbf{J}^{-1}, \quad (13)$$

where  $E[\cdot]$  determines the expectation value and  $\mathbf{p}_t$  is the true node position.

On the other hand, when dealing with a rather large number of gateway combinations, not all relative configurations with respect to the unknown node position are equivalent. Therefore, the proposed method invokes the paired CRLB information applied on each combination of 2 gateways to detect these most suitable hyperbolas associated with the best configurations. Thus, the node is localized after removing the combinations which have the highest paired CRLB values at an initial guess location  $\mathbf{p}_o = [x_o, y_o]^T$  rather than, using the whole combinations including the ones whose configurations are less favorable and have been observed to introduce large errors. This algorithm is achieved using the following two steps:

a) *Estimating the paired CRLB values:* is done separately for each combination by assuming a constant standard deviation  $\sigma$  for all the gateways. Thus, the covariance matrix at each pair of gateways is reduced to the scalar value  $R = 2\sigma^2$ . Accordingly, the Jacobian of the measurement function for a given pair of gateways  $i$  evaluated at position  $\mathbf{p}_o$  is calculated as:

$$\mathbf{f}_i^o = \left. \frac{\partial h_{l(i),r(i)}}{\partial \mathbf{p}} \right|_{\mathbf{p}=\mathbf{p}_o} \quad (14)$$

$$= \left[ \left. \frac{\partial h_{l(i),r(i)}}{\partial x} \right|_{\mathbf{p}=\mathbf{p}_o}, \left. \frac{\partial h_{l(i),r(i)}}{\partial y} \right|_{\mathbf{p}=\mathbf{p}_o} \right] \quad (15)$$

$$= \left[ \frac{x_o - x_{l(i)}}{d_{l(i)}^o} - \frac{x_o - x_{r(i)}}{d_{r(i)}^o}, \frac{y_o - y_{l(i)}}{d_{l(i)}^o} - \frac{y_o - y_{r(i)}}{d_{r(i)}^o} \right], \quad (16)$$

with  $d_{l(i)}^o = \|\mathbf{g}_{l(i)} - \mathbf{p}_o\|_2$  and  $d_{r(i)}^o = \|\mathbf{g}_{r(i)} - \mathbf{p}_o\|_2$ .

Hence, the computation of the Fisher Information Matrix follows as:

$$\mathbf{J}_i^o = \frac{1}{2\sigma^2} \mathbf{f}_i^{oT} \mathbf{f}_i^o, \quad (17)$$

thus, the paired CRLB value is calculated from the inverse of the FIM  $\mathbf{J}_i^o$  then, calculating the square root of its trace as:

$$c_i^o = \sqrt{\text{tr}(\mathbf{J}_i^{o-1})}, \quad (18)$$

b) *Determining the outliers:* from all the determined paired CRLB values  $\mathbf{C}^o = [c_1^o, \dots, c_H^o]$ , which are shown in Figure 1, by observing that the high peak at any combination index is considered as an outlier combination, while the low peaks are the proper combinations to be utilized for the localization. Thus, several thresholding methods have been investigated, which all try to separate the outliers from the regular paired CRLB values without other a priori knowledge. These investigated thresholding techniques are mainly based on analyzing the variance of the paired CRLB values since it is observed that the regular values have amplitudes close to one another. Subsequently, all these methods assume that the combinations have two classes: ordinary paired CRLB value class and outlier class which are below and above the shown threshold in Figure 1, respectively.

In this paper, the proposed thresholding technique is based on Median Absolute Deviation (MAD) which is a robust measure of how spread out a set of data is [11]. Hence, the classical standard deviation estimator could also measure the spread, but it is more affected by the extremely high or extremely low values and non-normality due to the presence of outliers. Accordingly, this threshold is computed as:

$$\gamma = \text{median}(\mathbf{C}^o) + 3 \cdot \hat{\sigma}_{MAD} \quad (19)$$

with

$$\hat{\sigma}_{MAD} = b \cdot \text{median}(|\mathbf{C}^o - \text{median}(\mathbf{C}^o)|), \quad (20)$$

where 3 has been heuristically chosen and the normalization factor  $b$  is set to 1.4826 in order to be consistent for the

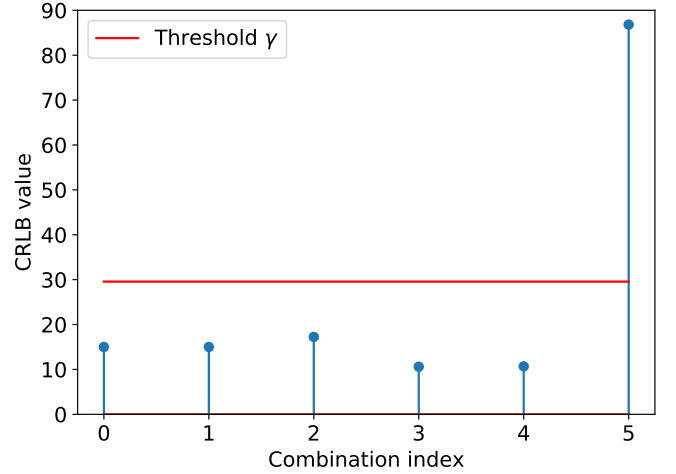


Fig. 1: The CRLB values of all the combinations ( $H = 6$ ).

standard deviation at the normal Gaussian distribution i.e., usually the region of interest.

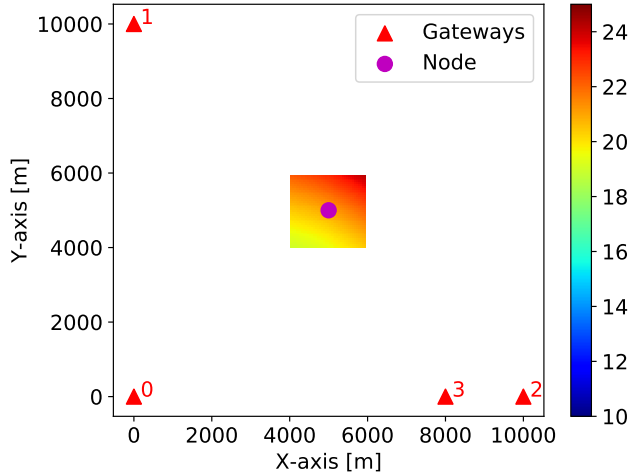
After labeling the outlier CRLB values by this threshold technique, the node is localized without these combinations. As shown in Figure 2b, the CRLB values are reduced in the vicinity of the node location after removing the outlier combinations in comparison to the scenario in Figure 2a without filtering. Such a strategy reduces the probability of the error in the estimated position.

#### IV. SIMULATION MODEL

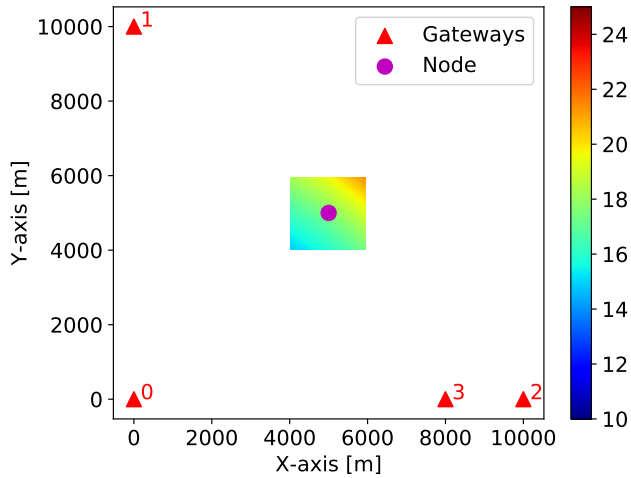
Two main aspects shape the simulator model which allows for a dynamic study of the system performance. These are the distributions of the gateway locations with respect to the node location and the choice of convenient uncertainty in the timestamps, to be closer to the real measurements as detailed in the following subsections.

##### A. Gateway locations

The gateway locations are produced by the Poisson disk distributions algorithm which has been introduced in [12]. This is considered as a fast 2-dimensional blue noise sampler, easily implemented in arbitrary dimensions and it is guaranteed to take  $O(M)$  time to generate  $M$  Poisson disk samples. To start the process, this algorithm takes as input the length  $len$  and width  $wid$  of the samples domain in  $\mathbb{R}^2$ , and the minimum distance  $\rho$  between the samples. First, it initializes a 2-dimensional background grid for storing samples and accelerating the spatial searches. Then, it selects the initial sample randomly chosen uniformly from the domain and inserts it into a cell in the background grid. The cell size is picked to be bounded by  $\frac{\rho}{\sqrt{2}}$ , so that each grid cell will contain at most one sample. In the next iteration, the neighboring point is chosen uniformly from the spherical annulus between radius  $\rho$  and  $2\rho$  around the previous sample. This linear algorithm is done recursively until all the  $M$  samples are generated. At this point, the node location is chosen randomly from the



(a) Without paired CRLB pre-processing.



(b) With paired CRLB pre-processing.

Fig. 2: CRLB plot using the full combinations.

samples, while maintaining  $N$  extracted samples to be the gateway positions. Notice that generating the node location out of the dense  $M$  samples set, forbid the node location to be less than  $\rho$  meter from any gateway.

#### B. Timestamp perturbation model

For the sake of simplicity, only the thermal noise using a normal Gaussian distribution is considered. Thus, the timestamp is individually calculated as in Equation 3 while assuming a zero mean of the uncertainty term  $u_n \sim \mathcal{N}(0, \sigma^2)$ .

### V. SIMULATION RESULTS

In the previous section, the simulation framework is provided including the location distributions and the chosen noise model. In this section, the simulation results are presented, using the Euclidean distance error between the true position of the node and the estimated location, as a measure of performance for the localization methods. This is measured

by Monte Carlo simulations after utilizing the parametric TDoA technique i.e., introduced in [13], with and without paired CRLB pre-processing for comparison. Moreover, the two main parameters which shape the simulation scenarios are the magnitude of the thermal noise variance and the number of gateways as detailed in the following subsections.

#### A. Impact of the noise variance

The robustness of the system against the outlier combinations is checked by changing the thermal noise standard deviation  $\sigma$  in Equation 3 to be in the range from  $0.1 \mu\text{s}$  to  $1 \mu\text{s}$  ( $\approx 30 \text{ m}$  to  $300 \text{ m}$ ), while maintaining the number of gateways to be  $N = 7$  which are chosen randomly from a fixed location distribution. For this gateway locations configuration, the length  $len$  and width  $wid$  of the map are fixed to  $5 \text{ km}$  and  $10 \text{ km}$ , respectively, while the minimum distance  $\rho$  between the  $M$  samples is fixed to  $100 \text{ m}$ .

As shown in Figure 3a, the simulation result shows a noticeable reduction in the localization error medians when using the paired CRLB pre-processing algorithm. Accordingly as shown in Figure 3b, the CDF curves obtained for all the noise values preserve the same performance rank over the whole simulations with 50% of the error values less than  $200 \text{ m}$  and  $230 \text{ m}$ , while turning the proposed algorithm ON and OFF, respectively. This indicates that the proposed method is robust to the high noise variances.

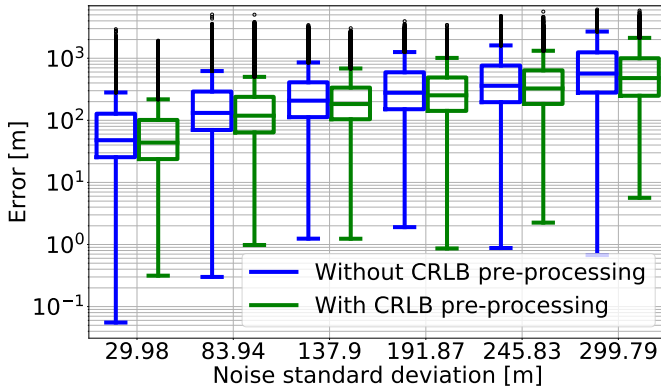
#### B. Impact of the number of gateways

The localization performance assessment for various number of gateways is studied by utilizing the same gateway locations configuration i.e., introduced in the previous subsection, while choosing randomly from them particular number of gateways  $N$  to be in the range from 3 to 15. Moreover in all the results, the thermal noise standard deviation  $\sigma$  is assumed to be equal to  $0.1 \mu\text{s}$  ( $\approx 30 \text{ m}$ ).

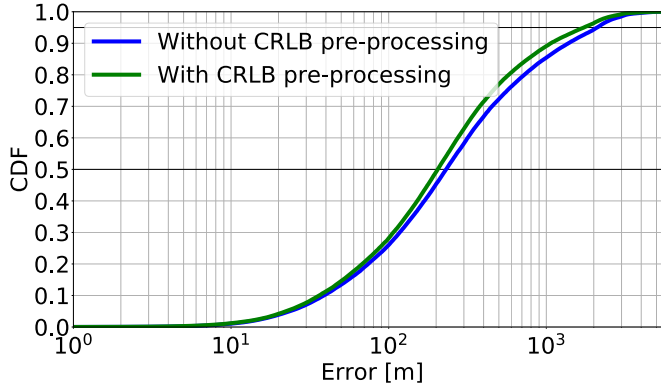
As shown in Figure 4a, it is clear that all the medians of the proposed method are drastically the lowest values for all the network densities, especially for the large number of gateways ( $> 3$ ) i.e., a realistic value in the near future. Thus, it is clear in the lower network density that the number of hyperbolas decreases ( $\approx 3$ ) accordingly, only one hyperbola might be dropped resulting in a tiny or almost no gain. Nevertheless as shown in Figure 4b, the CDF curves obtained for all number of gateways range still confirm the prevalence of the proposed method over most of the simulations with 95% of the error values less than  $1600 \text{ m}$  and  $1940 \text{ m}$ , while turning the proposed algorithm ON and OFF, respectively.

### VI. CONCLUSIONS

In this article, a novel TDoA pre-processing method for IoT localization is presented. This technique improves the accuracy of positioning in the presence of outlier hyperbolas by choosing only the best hyperbolic constraint sets for estimating the node location based on their low paired

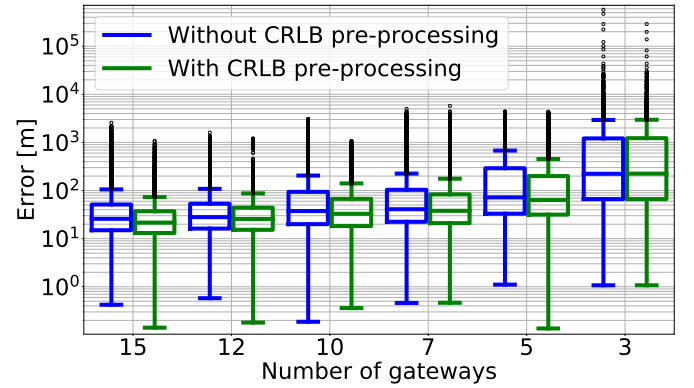


(a) Errors at each value of the noise standard deviation.

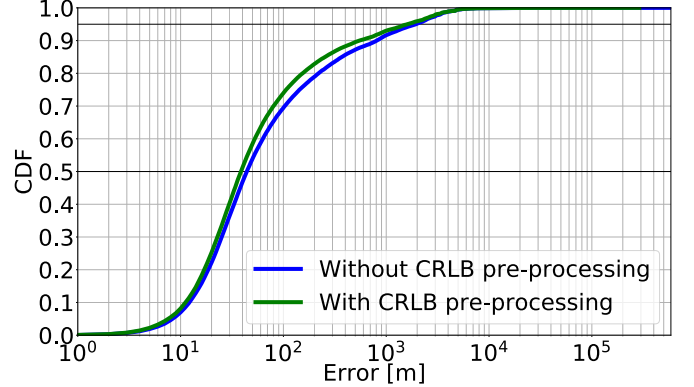


(b) The whole simulation results of  $\sigma$  ranging from 30 m to 300 m.

Fig. 3: Euclidean distance error and CDF while changing noise standard deviation  $\sigma$ .



(a) Errors at each number of gateways  $N$ .



(b) The whole simulation results of  $N$  ranging from 3 to 15.

Fig. 4: Euclidean distance error and CDF while changing the number of gateways  $N$ .

CRLB values. Thus, the principle is based on dropping the combinations, whose paired CRLB values are high in the proximity of the node location, using a robust thresholding technique. For performance assessment of this algorithm, a simulator has been developed. It uses a Poisson distribution approach parameterized for setting the location of gateways and nodes as close as possible of the realistic situations. The used simulator also emulates the proper disturbance in the timestamp values. Simulation results demonstrate the high performance of our pre-processing algorithm over a wide range of noise standard deviations and network densities.

In future work, an optimum technique for outliers detection could be investigated rather than using a thresholding method.

## REFERENCES

- [1] M. Centenaro, L. Vangelista, A. Zanella and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios," in *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60-67, October 2016.
- [2] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui and T. Watteyne, "Understanding the Limits of LoRaWAN," in *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34-40, Sept. 2017.
- [3] L. Mainetti, L. Patrono, A. Secco and I. Sergi, "An IoT-aware AAL system for elderly people," 2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech), Split, 2016, pp. 1-6.
- [4] A. H. Seyed, A. Tarighat and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," in *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24-40, July 2005.
- [5] Z. Li, T. Braun, X. Zhao, Z. Zhao, F. Hu and H. Liang, "A Narrow-Band Indoor Positioning System by Fusing Time and Received Signal Strength via Ensemble Learning," in *IEEE Access*, vol. 6, pp. 9936-9950, 2018.
- [6] Choi, Wongeun and Chang, Yoon-Seop and Jung, Yeonuk and Song, Junkeun. (2018). Low-Power LoRa Signal-Based Outdoor Positioning Using Fingerprint Algorithm. *ISPRS International Journal of Geo-Information*.
- [7] Zafari, F., Gkelias, A., and Leung, K. K. (2019). A Survey of Indoor Localization Systems and Technologies. *IEEE Communications Surveys and Tutorials*, 1-1.
- [8] B. C. Fargas and M. N. Petersen, "GPS-free geolocation using LoRa in low-power WANs," 2017 Global Internet of Things Summit (GloTS), Geneva, 2017, pp. 1-6.
- [9] Alan Bensky. 2007. *Wireless Positioning Technologies and Applications*. Artech House, Inc., Norwood, MA, USA.
- [10] R. Kaune, "Accuracy studies for TDOA and TOA localization," 2012 15th International Conference on Information Fusion, Singapore, 2012, pp. 408-415.
- [11] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," in *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61-80, July 2012.
- [12] Robert Bridson. 2007. Fast Poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH 2007 sketches (SIGGRAPH '07)*. ACM, New York, NY, USA, , Article 22 .
- [13] Ahmed Ghany, Bernard Uguen, Dominique Lemur. A Parametric TDoA Technique in the IoT Localization Context. *WPNC*, Oct 2019, Bremen, Germany. hal-02302985