Abstract—This work proposes and evaluates a hybrid Bayesian-based localization method to estimate the position of a target node using received signal strength and time of flight measurements. In our investigations, we consider these measurements are acquired through a distributed antenna system which is connected to a common master anchor node. The baseline non-hybrid scenarios use only received signal strength measurements to estimate the position of interest, while the hybrid implementation combines time of arrival measurements as well. Both Bayesian-based (non) hierarchical approaches approximate the posterior distribution of the target’s location coordinates using Markov Chain Monte Carlo methods. The hierarchical method introduces conditional interdependencies to the model parameters, resulting in less model variance. Herein, the root mean square error is used to evaluate the performance of the indoor test scenarios. Our results show that both hybrid and hierarchical approaches outperform the baseline Bayesian model, while the former significantly increase the accuracy the target position estimate.

I. INTRODUCTION

The 5G New Radio (NR) system will support exciting new applications which are organized into three use cases, namely massive Machine Type Communications (mMTC), the Ultra Reliable Low Latency (URLLC) and the extreme Mobile Broadband (eMBB) communications [1]. Moreover, owing to its large bandwidth, very high carrier frequency and massive antenna array, the 5G NR systems put forward a plethora of compelling new deployment scenarios which require or benefit from accurate and reliable position information [2].

The location information becomes particularly advantageous for emerging IoT (Internet of Things) applications across various industry verticals: for example, assets tracking, context-aware marketing, transportation and logistic systems, augmented reality, health care, as well as haptic technologies. In such indoor deployment scenarios, traditional localization methods relying on GPS (Global Positioning System) satellites and standard cell multilateration are limited or even impractical. And yet traditional location estimation based on time of flight suffers from synchronization mismatch in such infrastructureless mMTC deployments.

Academia and industry have motivated to overcome all aforementioned challenges, designing new techniques and algorithms suitable for Indoor Positioning Systems (IPS) due to its importance in several application areas [3]–[7].

To estimate the target node position, we can employ distinct measurements depending on the radio access interface, propagation features and intended accuracy of the localization strategy. An IPS can implement distinct computational methods to estimate the target position using one or a combination of the following alternatives: Time of Arrival (TOA) [8], Time Difference of Arrive (TDOA), Received Signal Strength (RSS) [9] and Direction of Arrival (DOA) [10].

Following a frequentist approach, traditional methods to find the target position resort to Nonlinear Least Squares (NLS) and Maximum Likelihood (ML) estimators [11]. Recently, Bayesian inference using probabilistic graphical models emerged as a promising alternative to estimate the location of the target node by sampling the joint posterior distribution of its position coordinates. In this context, Markov Chain Monte Carlo (MCMC) methods constitute a robust approach to sample large Bayesian models that, otherwise, would require multidimensional integration over several unknown parameters. In indoor deployment scenarios, such Bayesian graphical models can use distinct metrics, such as TOA, TDOA, RSS and DOA, or a hybrid approach (combining distinct metrics) to find the position of a target [1], [12].

Deployments of mMTC in low latency and high reliability scenarios are still a challenge. Communication systems has been revisited to minimize the utilization of resources and satisfy the requirements of the system. Recently, there has been considerable interest in physical layer techniques for those types of scenarios [13], [14]. Distributed Antenna System (DAS) have been approached in recent works. In [15] authors propose DAS to reduce the use of spectrum resources from the conventional co-located antenna system and satisfy the URLLC requirements. In [16], authors explore short packet length satisfying the low latency for industrial IoT, investigating the packet error probability in the DAS for downlink URLLC.

Our work proposes and evaluates Bayesian-based localization methods using RSS and TOA measurements in indoor deployment scenarios employing distributed antenna system. The reminder of this paper is organized as follows. Section II presents the Bayesian networks by introducing the probabilistic graphical models and the MCMC algorithms used to estimate the posterior distribution of the target position. In Section III, we discuss localization’s models and the evaluation scenarios under consideration. Section IV presents the numerical results in terms of the root mean square error (RMSE) for distinct Bayesian network layouts and input measurements. Finally, Section V concludes this work and provides final remarks.

II. BAYESIAN-BASED MODEL AND INFERENCE

Bayesian networks represent the conditional interdependencies between random variables in probabilistic graphical models while allowing to infer the likelihood of the related events conditional on the available prior knowledge. By using such networks, it is possible to predict the occurrence of events as well as update the prior knowledge through an iterative procedure when new information is acquired, thus, reducing the corresponding uncertainty.

A. Graphical Models

A graphical model is a multivariate statistical model that represents the conditional interdependencies between descriptive parameters. In such models, the vertices correspond to
random variables (RVs) while the edges represent the underlying relationships between them [17]. We model the source localization system by Directed Acyclic Graphs (DAGs) and infer the target node position. In this model, the RVs represented as vertices are assumed to be conditionally independent. In other words, a vertex is only affected by its own parents, while being independent of its non-descendants given its parents. This property implies a factorization of the joint probability density function \( p(V) \) of the RVs \( X_v, v \in V \), where \( V \) is a set of RVs, constituting the model as given next

\[
p(V) = \prod_{v \in V} p(v | \text{pa}(v)), \tag{1}
\]

where \( \text{pa}(v) \) represents the parents of \( v \). The conditional distribution of a RV \( v \) in the graph is given by

\[
p(v|V/v) \propto p(v, V/v) \\
= p(v|\text{pa}(v)) \prod_{w \in \text{child}(v)} p(w|\text{pa}(w)), \tag{2}
\]

where \( \text{child}(v) \) yields all the children of \( v \). In this work, we use Python (a general-purpose programming language) to implement a Bayesian-based source localization system. In particular, we employ the PyMC3 package [18] to build Bayesian statistical models and carry out approximate inference using the Markov Chain Monte Carlo (MCMC) approach. The MCMC approach uses the Bayes’ Theorem to estimate the prior knowledge.

B. Bayes’ Theorem and the Posterior Distribution

Considering the RVs and their conditional interdependencies, the Bayes theorem is used to estimate the posterior distribution of interest, based on the likelihood function and prior distributions. From [19], the posterior distribution can be estimated as follows,

\[
p(H|D) = \frac{p(D|H)p(H)}{p(D)}, \tag{3}
\]

where \( H \) is the hypothesis of the system and \( D \) is the observed data, \( p(H) \) is the prior distribution and represents the initial hypothesis about the system parameters, \( p(D|H) \) is the likelihood function and \( p(D) \) is the evidence that represents the probability of all possible values that the parameters can assume, it is used as a normalization factor.

The data parameters can be organized into groups to obtain a hierarchical structure, which results in conditional independencies between those parameters [20]. This strategy is usable when information is available on different levels of observational units. The model is called Hierarchical and is more accurate than the non-hierarchical models, because the posterior distribution is less sensitive to multilevel model data, resulting in less model variance than the classic method. A one-level hierarchy with hypothesis of the system \( H = (\theta, \phi) \) and data \( D \) [21] is given by

\[
p(\theta, \phi|D) \propto \prod_{i=1}^{n} p(D_i|\theta_i)p(\theta_i|\phi)p(\phi). \tag{4}
\]

To illustrate this construction, we consider the one-way normal model

\[
y_i \sim N(\theta_i, \sigma_i^2), \quad \theta_i \sim N(\mu, \tau^2), \quad i = 1, \ldots, I,
\]

which relates to (4) by considering \( D = (y_i, \theta_i) \) and \( \phi = (\mu, \tau) \). In this case \( y_i \) is an observation and \( \theta_i \) is a parameter governing the data generation process for \( y_i \). The parameters \( \theta_1, \theta_2, \ldots, \theta_i \) are generated from a common population, with distribution given by a hyperparameter \( \phi \). Hence, the posterior distribution of this Hierarchical Bayesian Model is given by,

\[
p(\theta, \phi|D) = \frac{p(D|\theta, \phi)p(\theta|\phi)p(\phi)}{p(D)}, \tag{5}
\]

In this work, we build such (non) hierarchical models and use the Bayes’ Theorem through the MCMC approach to estimate the posterior distribution of the target node coordinates.

C. The Markov Chain Monte Carlo algorithm

The MCMC algorithms generate a sequence of random samples from a probability distribution and present a good performance when computing complex posterior distributions on high dimensional models. The Metropolis-Hastings algorithm implements the MCMC method to sample from a target distribution by first making a random proposal for new parameter values and then accepting or rejecting the proposal. From [22], a sample \( h^* \) drawn from target distribution must satisfy the following equation to be accepted,

\[
A_k(h^*, h^{(\tau)}) = \min \left( 1, \frac{\tilde{p}(h^*)q_k(h^{(\tau)}|h^*)}{\tilde{p}(h^{(\tau)})q_k(h^*|h^{(\tau)})} \right), \tag{6}
\]

where \( k \) yields the members of the set of possible transitions being considered, \( \tau \) is a step of the algorithm, \( h^{(\tau)} \) is the current state containing all the draw samples, \( h^* \) is a draw sample analysed, \( q_k(h|h^{(\tau)}) \) is the distribution of \( h^* \), \( A_k(h^*, h^{(\tau)}) \) is the probability for acceptance, and \( \tilde{p}(h) \) is a distribution that evaluate \( p(h) \) easily by the following equality \( p(h) = \frac{\tilde{p}(h)}{\tau_k} \) and \( H_{xy} \) is unknown normalizing constant. If the draw sample is accepted, then \( h^{(\tau+1)} = h^* \), otherwise the candidate sample \( h^* \) is rejected, \( h^{(\tau+1)} \) is set to \( h^* \) and another candidate sample is drawn from the distribution \( q_k(h|h^{(\tau+1)}) \).

The Metropolis-Hastings algorithm is based on a random walk (stochastic process) and may linger (re)-exploring the same region of the sample space for a number of steps, thus wasting computational resources. To cope with this situation, the No-U-Turn Sampler (NUTS) algorithm, based on the Metropolis-Hastings, is used to find a good estimates adaptively. Furthermore, it avoids re-exploring local spaces to ensure a much shorter simulation time [23].

III. SYSTEM MODEL

A. Evaluation Scenario

In this contribution, we evaluate the performance of the proposed solution in an indoor deployment scenario with a square shape and an anchor (central processing unit) with distributed antenna heads in known positions (in the corners of the area). The central unit then estimates the position of a target node after acquiring a minimum amount of measurements from the antenna heads. A target is randomly located inside this square and measurements are collected by the anchors. Each side of this square scenario has 100 meters as shown in Fig.1.

The radio links are assumed to be affected by a log-distance shadowed path loss model, so that each receiver collects the corresponding Received Signal Strength (RSS) measurements independently. Then, it is possible to determine the unknown position of the target node in this evaluation scenario by using the measurements gathered by at least 3 anchor receivers [24]. The radio link RSS function is given by

\[
p_{r_i} = p_t - \alpha \ln(d_i) + \varphi \text{ in dBm}, \tag{7}
\]
where $p_r$, is the received signal strength at the $i$th receiver, $p_t$ is the transmitted signal strength, $\alpha$ is the path loss exponent, $d_i$ is the euclidean distance between the target and the $i$th receiver, and $\varphi$ is the shadowing value with zero-mean normal distribution in logarithmic scale.

In the hybrid scenarios, two arbitrary anchors are assumed to collect RSS and TOA measurements. The TOA function is

$$ t_i = \frac{d_i}{c}, $$ \hspace{1cm} (8)

where $c$ is the speed of light and $t_i$ is the time of flight assuming that the source emits a signal at time 0 and the sensor receives it at time $t_i$. It is worth mentioning, assuming a distributed antennas system, the central unit is responsible for making the necessary synchronization between the anchors to acquire this type of data.

### B. Localization Graphical Models

We consider three types of graphical models: i) a RSS-based Bayesian network; ii) a hybrid model with the combination of RSS and TOA measurements; and iii) a RSS-based hierarchical Bayesian network. In the following we present the Directed Acyclic Graphs (DAG) model and respective metrics considered for each case.

i) **RSS-based Bayesian network:** The interdependence between the random variables in the RSS-based Bayesian network is represented by the DAG model shown in Fig.2. The illustration describes the following variables: $x$ and $y$ represent the receivers’ coordinates, which are assumed to be known a priori by the central unit; and remaining parameters, shown inside the circles, correspond to random variables whose distribution is set based on our prior knowledge. The assumptions of the respective problem are given as

- $p_t \sim \text{Normal}(0, 100)$, $\alpha_i \sim \text{Normal}(0, 100)$,
- $\sigma_i^2 \sim \text{Halfnormal}(10)$, $\mu_i \sim p_t - \alpha \log(d_i)$,
- $X \sim \text{Uniform}(0, L)$, $Y \sim \text{Uniform}(0, B)$,
- $d_i \sim \sqrt{(X - x_i)^2 + (Y - y_i)^2}$,

where $p_t$ is the transmit power, $\alpha$ is the path loss exponent, $\sigma_i$ is the standard deviation of the measurements collected by the $i$th receiver point, and the $i$th receiver and $(X, Y)$ are the RVs representing the target location coordinates.

ii) **Hybrid-based Bayesian network:** The hybrid model is a combination between a RSS- and TOA-based Bayesian networks. The interdependence between the random variables in latter model, using time of flight measurements, is represented by the DAG in Fig.3. Similarly to the RSS-based model, the TOA-based Bayesian’s network is described by

- $\mu_i \sim \frac{d_i}{c}$,
- $\sigma_i^2 \sim \text{Halfnormal}(10/c)$,
- $X \sim \text{Uniform}(0, L)$, $Y \sim \text{Uniform}(0, B)$,
- $d_i \sim \sqrt{(X - x_i)^2 + (Y - y_i)^2}$.

The employed hybrid model combines the components presented in Figs. 2 and 3.

iii) **RSS-based hierarchical Bayesian network:** in this configuration, we create an additional one-level hierarchy to two data parameters, so that now the transmit power and the path loss exponent have a distribution with mean and variance given by $(a_i, \sigma^2_{a_i})$ and $(b_i, \sigma^2_{b_i})$, respectively, for each parameter. Comparing this model to i), the hierarchical structure introduces conditional independencies in these parameters resulting in less model variance. The DAG model shown in Fig. 5 illustrates the interdependence between the random variables in this network.

The diagram illustrates the relationships and conditional independencies among the variables in the hierarchical Bayesian network.
Fig. 4. The probability density of the source location estimate with (a) 20 and (b) 50 RSS measurements, and (c) 20 TOA measurements. In each plot, the real location of target is denoted by red circle denotes and the maximum of the probability density by black circle.

Fig. 5. Graphical model of the hierarchical RSS-based source localization.

Similar to the previous models, the hierarchical Bayesian network is described as

\[
p_{t_i} \sim \text{Normal}(a_i, \sigma_{a_i}^2), \quad a_i \sim \text{Normal}(0, 10),
\]

\[
\sigma_{a_i}^2 \sim \text{Gamma}(5, 1), \quad \alpha_i \sim \text{Normal}(0, \sigma_{b_i}^2),
\]

\[
b_i \sim \text{Normal}(0, 10), \quad \sigma_{b_i}^2 \sim \text{Gamma}(5, 1),
\]

\[
\sigma^2 \sim \text{Halfnormal}(10), \quad \mu_i \sim p_{t_i} - \alpha \log(d_i),
\]

\[
X \sim \text{Uniform}(0, L), \quad Y \sim \text{Uniform}(0, B),
\]

\[
d_i \sim \sqrt{(X - x_i)^2 + (Y - y_i)^2},
\]

where \(a_i\) and \(b_i\) are assumed to be a normal distribution with mean 0 and with variance 10 and represent the mean of \(p_{t_i}\) and \(\alpha_i\), respectively, \(\sigma_{a_i}^2\) and \(\sigma_{b_i}^2\) represent the variance of \(p_{t_i}\) and \(\alpha_i\) and are assumed to be a gamma distribution with shape parameter 5 and a scale parameter 1.

IV. PERFORMANCE EVALUATION

We evaluate the Bayesian-based source localization methods for indoor deployment scenarios (see § III). We implement our estimator using the PyMC3 Python package and carry out an extensive simulation campaigns to assess the proposed solutions. In our investigations, the target location is unknown by the algorithm and measurements are generated according to the radio propagation channel described in (7) and (8). Then, anchors collect the measurements and forward them to the central unit which employs the NUTS algorithm to sample the posterior distribution of a bidimensional coordinate representing the target position \((X, Y)\). The target node is located at \((70, 30)\) in every simulation scenario. We ran simulations by varying the number of measurements collected by each receiver and fed to the NUTS algorithm. The anchors independently acquire the measurements, the real target position is represented by the red circle, the estimated target position is the black point in the PDF illustrated by blue color. The results for RSS- and TOA-based Bayesian network are presented in Fig.4.

Our results show that the number of measurement samples acquired by the receivers affects the accuracy of the algorithm. In fact, by comparing the RSS-based approach with 20 and 50 samples in Figs. 4(a) and 4(b) respectively, the latter shows a more accurate estimate of the target position. Thus, we conclude that RSS-based approach performs better by acquiring more measurements (more evidence). In Fig. 4(c), we observe that the TOA-based approach needs much less measurements to achieve a similarly performance.

Fig. 6 shows the accuracy improvement with the use of hybrid models. In this simulation set, we compare the Root Mean Square Error (RMSE) between the distinct models by varying the standard deviation of the error associated to the

Fig. 6. Root mean square error for increasing standard deviation when using of RSS and hybrid (RSS plus TOA) models.
RSS measurements acquired by the receivers. Each point in the curve is averaged over 60 simulation runs and in each such iteration the localization algorithm is fed with 20 new RSS measurements – notice that the hybrid approach uses new RSS and TOA measurements for each run. Anchors independently acquire RSS or TOA measurements and the NUTS algorithm is used to estimate the position. Similar to the results in Fig. 4(c), the hybrid approach outperforms the RSS-based one, though the performance gap narrows by increasing the standard deviation of the RSS measurements error. This occurs because the hybrid approach has more information acquired by TOA measurements and this compensates the high values of standard deviation of the RSS measurements error for estimating the location of the target node.

In a similar way, Fig. 7 compares the hierarchical approach (see § II-B) against the RSS-based models in terms of the RMSE as function of standard deviation of the error associated with the measurements. In this figure, each point in the curve is averaged over 60 simulation runs and in each such iteration the localization algorithm is fed with 20 new RSS measurements. Anchors independently acquire RSS measurements and the NUTS algorithm is used to estimate the position. The hierarchical models achieve better performance than non-hierarchical models. Furthermore the gap in performance becomes higher when the error associated with the measurements increase. This occurs because the hybrid model consider conditional independencies between the parameters of the model and this result in less model variance than the non-hierarchical method to estimate the target position.

V. CONCLUSIONS

In this contribution, we evaluate three Bayesian-based localization methods to estimate the position of a target node. The network represents a distributed antenna system connected to a central unit that synchronizes the different types of acquired data. The NUTS algorithm is used to sample the posterior distribution and estimate the target position. The results show that hybrid Bayesian-based method can be efficient to estimate the localization node and outperforms the non-hybrid Bayesian-based method. Our results also show that hierarchical models perform better in scenarios when the error associated with the measurements increase. A combination of the hybrid Bayesian-based method with the hierarchical approach is the intention of our next work. That is expected to provide more accuracy in localization.

ACKNOWLEDGMENT

The research leading to these results has received funding from the Academy of Finland through the grants No. 318927 (project 6Genesis Flagship) and No. 24303208.

REFERENCES