On the Performance of AoA–Based Localization in 5G Ultra–Dense Networks

ESTIFANOS YOHANNES MENTA1, NICOLAS MALM1, RIKU JÄNTTI1 (Senior Member, IEEE), KALLE RUTTIK1, MÁRIO COSTA2 (Member, IEEE), AND KARI LEPPÄNEN2

1Department of Communications and Networking, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland
2Radio Network Technologies Team, Huawei Technologies Oy (Finland) Co. Ltd., 00180 Helsinki, Finland

Corresponding author: Estifanos Yohannes Menta (e-mail: estifanos.menta@aalto.fi)

This work was supported in part by the 5th Evolution Take of Wireless Communication Networks (TAKE-5), Finnish National 5G program through TEKES (Finnish National Foundation), and in part by the PriMO-5G project through the European Union’s Horizon 2020 Research and Innovation Programme under Grant 815191.

ABSTRACT
Cellular systems are undergoing a transformation toward the fifth generation (5G). Envisioned applications in 5G include intelligent transport system (ITS), autonomous vehicles, and robots as a part of future roads, factories, and society. These applications rely to a great extent on accurate and timely location information of connected devices. This paper proposes a practical scheme for acquiring precise and timely position information by means of an edge cloud. The considered solution consists of estimating and tracking the azimuth angle–of–arrival (AoA) of the line–of–sight (LoS)–path between a device and multiple transmission–reception points (TRPs), each having a uniform linear antenna array (ULA). AoA estimates from multiple TRPs are fused into position estimates at the edge cloud to obtain timely position information. The extensive measurements have been carried out using a proof–of–concept software–defined–radio (SDR) testbed in order to experimentally assess the achievable positioning accuracy of the proposed architecture. A realistic UDN deployment scenario has been considered in which TRPs consist of antenna arrays mounted on lamp posts. Our results show that practical UDNs can provide sub–meter positioning accuracy of mobile users by employing ULAs with at least four antennas per TRP and by taking into account the non–idealities of the ULAs’ phase and magnitude response.

INDEX TERMS
AoA, localization, position, UDN, edge cloud.

I. INTRODUCTION
The ongoing transformation of cellular systems towards future 5G networks, with support for machine–to–machine (M2M) communication and internet of things (IoT) services, is driving new markets and industry segments beyond the traditional human–centric communication. Emerging applications of 5G, which are envisioned to be part of future roads and factories, include autonomous vehicles, autonomous robots and ITS use–cases that enable cooperative collision avoidance, high density platooning and vulnerable road user discovery [1], [2]. The majority of these 5G use–cases are dependent on precise and timely location information of mobile devices. As a result, providing position information has changed from an on–demand approach with coarse position estimation in the past to an always–on approach in 5G with strict requirements in the level of accuracy, availability, continuity and integrity as well as trust in the solution. Location awareness is becoming an essential element of many emerging markets, and positioning is considered an integral part of the system design of upcoming 5G [3], [4].

Various radio–based positioning methods have been studied in the literature. One such method of obtaining location information is based on the received signal strength (RSS) of radio signals exchanged among base stations (BSs) and mobile devices [5]–[8]. Positioning may also be obtained by means of time–of–arrival (ToA)–based methods where the distance between transmitter and receiver is found from the propagation delay and propagation speed, given that the transmission time is known. In case such an assumption does not hold true, differences of reception time at different locations, known as time–difference–of–arrival (TDOA) [9]–[12], may be exploited. A third approach consists in using an array of antennas and exploiting the AoA of the LoS path between...
every user equipment (UE) and each BS [13]. Then, the AoA estimates acquired by all BSs are fused at a central unit in order to determine the devices’ locations; see [3], [14] and references therein.

The strict requirements on latency and accuracy of position information for real–time services of 5G is well beyond that achieved by existing legacy systems, e.g., long–term evolution (LTE) or Global Navigation Satellite Systems (GNSSs). This calls for a solution that allows for accurate positioning with ultra–low–latency in order to satisfy the requirements of emerging use–cases.

One of the enablers of 5G is UDNs. UDNs facilitate accurate positioning due to a higher likelihood of having LoS communication most of the time, both in indoor and outdoor scenarios. Additionally, a device in a UDN can be connected to more than one small BS, commonly termed TRP. Each TRP provides an estimate of spatial and temporal parameters, based on its measurements, to a central unit for computation of position estimates. UDNs have also the potential of boosting area capacity and enhancing the end–user experience by overcoming cell boundary problems such as inter–cell interference (ICI), signalling overhead in mobility management and frequent handover (HO) by employing a so–called user–centric UDN approach [15], [16]. In a user–centric UDN, a UE is connected to multiple TRPs and the network decides which TRP is, or group of TRPs are, best suited to serve a given UE.

Performing localization in a fully–centralized architecture, typically located far away from the majority of UEs, is challenging due to the need for delivering timely location information under the strict limits imposed by 5G low–latency services. It is therefore necessary to perform localization in a distributed manner utilizing high–performance machines with adequate computing and storage capacity closer to the users for latency constrained services. To this end, edge computing is an emerging and effective approach for services requiring complex data processing and low–latency communication [17]. With edge clouds, processing and storage resources are made available at close proximity to the mobile terminals in the network. Therefore, edge cloud is responsible for providing precise and timely position estimation as well as user–centric services. An intelligent network having this controller at the edge would be responsible for joint optimization such as dynamically assigning a group of TRPs to serve each UE seamlessly, performing scheduling decisions, proactive handover and mobility management [18].

An important prerequisite for antenna array–based AoA estimation is phase–coherence. A time–invariant phase difference among the signals at the receiver chains is typically a requirement for AoA estimation of a target at specific location [14], [19]. However, in practical systems, with off–the–shelf radio front–end hardware, the relative phase among receiver chains varies with time due to drifts in the clocks driving each radio frequency (RF)–chain. Moreover, antenna arrays built in practice are typically subject to mutual–coupling, individual antenna beampattern and phase–center. Employing the theoretical array steering vector model in [20] for AoA estimation on a practical system typically leads to a significant performance degradation [21]. This calls for calibration techniques in order to compensate for time–varying phase–offsets as well as to take into account other gain and phase uncertainties that are unavoidable in any real–world system.

Motivated by the aforementioned practical limitations of state–of–the–art localization schemes, with regards to the requirements of envisioned use–cases for 5G, this paper contains the following contributions:

- An edge cloud–based architecture for real–time positioning services in 5G, requiring strict latency bounds and high accuracy, by fusing directional parameter estimates from each TRP comprising a UDN. The proposed architecture takes advantage of: 1) very small propagation delay in UDNs (≤ 83.33 ns in cells with inter–site distance (ISD) ≤ 50 m), 2) very small transmission delay due to very short slot duration design in the utilized frame structure. 5G New Radio (NR) has flexible frame structure with scalable subcarrier spacing (f_s) of 2^µ * 15 kHz, where µ ∈ {0, 1, 2, 3, 4, 5} and for µ ≥ 2, the slot duration ≤ 0.250 ms which is a suitable choice for low–latency services and 3) high–performance machines to significantly reduce the processing delay as a result of computationally demanding solutions, e.g. AoA estimation and localization.
- Validation of the proposed solution using a proof–of–concept SDR–based testbed developed at Aalto University. Performance measurements have been carried out in a realistic UDN scenario with TRPs having ULA antennas deployed on lamp posts.
- Identification of the usable azimuth angle range for accurate (sub–meter) positioning using ULAs. Such identified usable angle range can be used for designing future UDNs, namely for road users on highways.

The remainder of this paper is organized as follows. Section II provides the employed system model. Section III describes analysis of the considered AoA and position estimation techniques. Description of the experimental setup and testbed are given in Section IV. Numerical results obtained from testbed measurements are given in Section V. Finally, Section VI concludes the paper.

II. SYSTEM MODEL
We consider an edge cloud–based user–centric UDN positioning architecture in which small cells equipped with antenna arrays are deployed on lamp posts along the street; see Fig. 1. We focus on ULAs due to practical limitations of our testbed; see Section IV. In particular, our testbed could only support up to four antennas on each TRP. In such cases, linear arrays have a larger aperture than e.g. circular arrays when incident signal coming from source is around broadside as indicated in [22]. But the cost of ULAs is twofold: only a single direction (e.g., azimuth angle) can be unambiguously
which has 52 useful subcarriers. Once the UE is synchronized, LTE is adopted here as PSS to fit the utilized frame structure frequency–domain Zadoff–Chu (ZC) sequence [24] used in similar fashion to LTE. However, a truncated version of transmitted by TRP to identify frame and sample timings, to TRP by decoding Primary Synchronization Signals (PSS) synchronization and transmission of pilot signal for position-

grid allocation of the utilized frame structure can be found in current implementation and 5G NR is on the values of a 5G NR specification compliant one. The main differences had started before the 3GPP has frozen the 5G NR air interface parameters used. The detail on time–frequency domain allocation of the utilized frame structure has slot duration (also called transmission time interval (TTI)) of 0.184 ms, which falls in between \( \mu = \{ 2, 3 \} \) of NR. See also Table 1 for further details on testbed air interface parameters. Since testbed development has started before the 3GPP has frozen the 5G NR air interface specifications, the utilized radio frame structure is not a 5G NR specification compliant one. The main differences in current implementation and 5G NR is on the values of air interface parameters used. The detail on time–frequency grid allocation of the utilized frame structure can be found in [23]. Furthermore, extension of the current platform to 5G NR specification compliant one is ongoing separately and taken as future work.

A device, with a single antenna, connects to a synchronized set of TRPs by means of a network entry procedure to establish a communication link with the TRPs. This entails device synchronization and transmission of pilot signal for position-

FIGURE 1. Proposed architecture for positioning with an edge cloud–based user–centric UDN.

with TRP, it starts transmitting UL PSP which is generated in similar fashion from ZC sequence but with a different root index. The TRPs exploit this periodically transmitted OFDM–based PSP from UEs in uplink for AoA estimation. The frequency domain representation of the received signal from the antenna array can be written as:

\[
    r = SB_p(\phi, \tau)\xi + n, \tag{1}
\]

where \( S \in \mathbb{C}^{NN_f \times NN_f} \) denotes a diagonal matrix containing the PSP and \( \xi \in \mathbb{C}^2 \) is a vector containing the complex channel coefficients of the dominant (typically LoS) signal path. Moreover, \( B_p(\phi, \tau) = [b_{ph}(\phi, \tau), b_{pv}(\phi, \tau)] \), where \( b_{ph}(\phi, \tau) \in \mathbb{C}^{NN_f \times 1} \) and \( b_{pv}(\phi, \tau) \in \mathbb{C}^{NN_f \times 1} \) denote response vectors for horizontal and vertical polarization, respectively. The channel parameters \( \phi \) and \( \tau \) denote azimuth angle and time delay respectively. \( n \in \mathbb{C}^{NN_f \times 1} \) denotes the zero–mean complex–circular Gaussian distributed receiver noise, \( N_f \) is the number of sub–carriers comprising the OFDM waveform and \( N \) is the number of antennas in TRP.

The estimate of the channel response at a given TRP is obtained from \( r \) as follows:

\[
    \hat{h} = S^{-1}r
    = B_p(\phi, \tau)\xi + \hat{n}, \tag{2}
\]

where \( \hat{n} = S^{-1}n \). Here, superscript \( -1 \) denotes matrix inverse. The channel parameters \( \phi \) and \( \tau \) (measured in second) are estimated from \( \hat{h} \).

Phase coherence of multiple RF chains of transceivers can be achieved by coupling them with a common reference, e.g, the 10 MHz reference signal. However, a closer inspection of the instantaneous differential phase of the RF signals shows instability due to: 1) phase noise of the synthesizers at transceiver chains; 2) weak coupling at the 10 MHz, and a long synthesis chain up to the RF output; 3) temperature differences which cause a change in the effective electrical length of some synthesizer components. Although a 10 MHz common reference signal is used, the phase coherence at base station level cannot be maintained perfectly due to the dominance of the second factor mentioned above and also difference in circuitry constituting each radio frontend.

Another level of phase calibration is required, on top of the 10 MHz common reference, in order to compensate the dynamic phase offset due to the variation in clocks driving each RF chain. In this work, such a phase calibration is achieved by having a transmitter at a known position sending reference signals. Antennas’ feeding cables, connectors and other RF components have also been calibrated using a vector network analyzer (VNA).

Each practical antenna array element has an individual gain and phase characteristic. In addition, mutual coupling, cross–polarization effects and mounting platform reflections may be significant and should be taken into account by the AoA estimation algorithm. In order to take into account these non–idealties of real–world antenna arrays, it is important to incorporate the antenna arrays’ complex–valued response...
for all azimuth–angles into the AoAs estimation algorithm. This is done by measuring the antenna arrays’ responses in an anechoic chamber and finding the so–called Effective Aperture Distribution Function (EADF) [21].

In particular, the complex–valued multicarrier polarimetric beampattern of a real world antenna array can be expressed in terms of the EADF, denoted by \( G \), as follows [25]:

\[
B_p(\varphi, \tau) = \left[ b_{h}(\varphi) \otimes b_{f}(\tau) \right], \quad (3)
\]

where

\[
b_{h}(\varphi) = G_{h}d(\varphi) \quad (4a)
\]
\[
b_{y}(\varphi) = G_{y}d(\varphi) \quad (4b)
\]
\[
b_{f}(\tau) = G_{f}d(\tau). \quad (4c)
\]

In (3), \( \otimes \) denotes the Kronecker product. Moreover, \( b_{h}(\varphi) \in \mathbb{C}^N \) and \( b_{y}(\varphi) \in \mathbb{C}^N \) denote the array responses corresponding to horizontal and vertical polarizations, respectively. The corresponding EADFs are given by \( G = [G_{h}, G_{y}] \) and \( G_{f} \in \mathbb{C}^{N_y \times N_y} \) denotes the frequency–response of the receiver. The phase vector \( d(\varphi) \in \mathbb{C}^M \) is given by:

\[
d(\varphi) = \left[ e^{-j(k_1 \varphi)} \ldots e^{-j(k_M \varphi)} \right]^T, \quad (5)
\]

where \( M \) denotes the number of modes employed for representing the array response. The phase vector and \( b_{f}(\tau) \in \mathbb{C}^{N_y \times 1} \) is expressed similarly to (5) with \( \varphi \) and \( M \) replaced by \( 2\pi \tau f_s \) and \( N_f \), respectively.

### III. ANALYSIS OF ANGLE AND POSITION ESTIMATION

Once the beampattern is modelled effectively, localization is performed by adopting a two–stage extended Kalman filter (EKF)–based positioning engine in [3] and [14]. Here, general system dynamics and measurement model in (2) are non–linear and the EKF linearizes, using the Taylor series expansion, about an estimate of the current mean and covariance [26, Ch.13]. In the first stage, the local EKF–based computing engine at each TRP does the AoA estimation and tracking by rejecting outliers in AoA estimation. In the 2nd stage, a global EKF–based computing engine at the edge cloud performs fusion of AoA estimates from each TRP to obtain the position estimate from noisy measurement at each stage.

#### A. LOCAL EKF–BASED COMPUTING ENGINE AT TRP

Considering the azimuth angle \( \varphi \) as the parameter of interest in the considered positioning scheme, it is enough to formulate the EKF such that it only tracks the AoA and ToA. Formulation of EKF with useful parameters, i.e., concentrating out the nuisance parameters from the log–likelihood function and leaving a concentrated log–likelihood function only as a function of the parameters of interest, is computationally efficient approach. In the course of tracking the azimuth angle \( \varphi \) and ToA \( \tau \) using the EKF, a continuous white noise acceleration (CUMA) model is employed for the state evolution. Hence, the state–vector in the local computing engine of the \( i^{th} \) TRP at time–instant \( k \), denoted by \( \theta_k^i \in \mathbb{R}^4 \), can be written as:

\[
\theta_k^i = \begin{bmatrix} \tau_k^i \quad \varphi_k^i \quad \Delta \tau_k^i \quad \Delta \varphi_k^i \end{bmatrix}^T. \quad (6)
\]

Following the linear state evolution model that stems from the assumed CWNA model, the system evolves from state \( k \) to \( k+1 \) with predicted state–vector \( (\theta_k^{i\rightarrow}) \) and covariance of state–vector estimate \( (\xi_k^i) \) respectively as:

\[
\begin{align*}
\theta_k^{i\rightarrow} &= F \theta_{k-1}^{i\rightarrow} \quad (7a) \\
\xi_k^i &= F \xi_{k-1}^{i\rightarrow} F^T + Q \quad (7b)
\end{align*}
\]

where \( F \in \mathbb{R}^{4 \times 4} \) denotes the state transition matrix and \( Q \in \mathbb{R}^{4 \times 4} \) denotes the state noise covariance matrix, which are given by:

\[
F = \begin{bmatrix} I_2 & \Delta \tau I_2 \ 
\end{bmatrix}, \quad (8a)
\]
\[
Q = \begin{bmatrix} \sigma_{w\Delta \tau}^2 I_2 & \sigma_{w\Delta \tau}^2 I_2 \\
\sigma_{w\Delta \tau}^2 I_2 & 2 \sigma_{w\Delta \tau}^2 I_2 \\
\sigma_{w\Delta \tau}^2 I_2 & 2 \sigma_{w\Delta \tau}^2 I_2 \
\end{bmatrix}. \quad (8b)
\]

Here, \( \Delta \tau \) denotes the time–interval from state \( k \) to \( k+1 \) and \( \sigma_{w\Delta \tau}^2 \in \mathbb{R} \) denotes the state noise variance, which is a design parameter.

After performing channel estimation from PSP, the posterior mean state–vector estimate \( (\hat{\theta}_k^i) \) and covariance estimate \( (\xi_k^i) \) in the update step of the EKF are given as:

\[
\begin{align*}
\hat{\theta}_k^i &= \hat{\theta}_k^{i\rightarrow} + \Xi (\theta_k^{i\rightarrow})^{-1} \quad (9a) \\
\xi_k^i &= \hat{\xi}_k^i + \nu (\theta_k^{i\rightarrow}) \quad (9b)
\end{align*}
\]

where \( \Xi (\theta_k^i) \in \mathbb{R}^{4 \times 4} \) and \( \nu (\theta_k^i) \in \mathbb{R}^{4 \times 1} \) denote the observed Fisher information matrix (FIM) and score–function of the state evaluated at \( \theta_k^i \), respectively. The posterior mean state and covariance of the state weight as in [20]:

\[
\begin{align*}
\Xi (\theta_k^i) &= \frac{2}{\sigma_w^2} \left\{ \left( \frac{\partial \gamma}{\partial \theta_k^i} \right) \left( \frac{\partial \gamma}{\partial \theta_k^i} \right)^\dagger \right\} \quad (10a) \\
\nu (\theta_k^i) &= \frac{2}{\sigma_w^2} \left\{ \left( \frac{\partial \gamma}{\partial \theta_k^i} \right)^\dagger \gamma \right\} \quad (10b)
\end{align*}
\]

where \( \gamma \) is expressed as:

\[
\gamma = (I - B_p(\varphi, \tau) P_p^c(\varphi, \tau) \hat{h}). \quad (11)
\]

Here, \( \{ \} \) denotes real part of a complex quantity, supercripts \( ^\dagger \) and \( ^+ \) denote the conjugate transpose and Moore–Penrose pseudo–inverse respectively. Initialization of the EKF, i.e. initial state estimate \( (\tau_0^i, \varphi_0^i) \) is obtained by employing the space–time conventional beamformer, which is identical to the deterministic maximum likelihood estimator (MLE) for a single propagation path as in [14] and [25].
The initial covariance estimate \( \left( \xi_k^i \right) \) can be found by evaluating the inverse of the observed FIM at \( (t'_i, \phi'_i) \). Also, the corresponding initial rate-of-change of these parameters as well as the update of the corresponding covariance matrix are carried out once two EKF estimates of the state have been obtained. See Algorithm 1 for further detail on implementation of initial estimation and then tracking of \( \tau \) and \( \phi \).

B. GLOBAL EKF–BASED POSITIONING ENGINE AT EDGE CLOUD

An edge–cloud positioning engine with prior location information of each TRP performs localization of a device by fusing the AoA estimates from multiple TRPs. The initial position estimate can be carried out by either a least–squares (LS) or total least–squares (TLS) method. The EKF at the edge–cloud tracks the 2D position \((x_k \text{ and } y_k)\) and 2D velocity vector \((v_{x_k} \text{ and } v_{y_k})\) of the user. Hence, the state–vector in the edge computing engine at time–instant \( k \), denoted by \( p \in \mathbb{R}^4 \), is given by:

\[
P_k = \begin{bmatrix} x_k & y_k & v_{x_k} & v_{y_k} \end{bmatrix}^T.
\] (12)

Assuming that the velocity of the UE is nearly constant during the transition from state \( k - 1 \) to \( k \), except only being perturbed by small random changes, the CWNA model still holds for state evolution, and thus the prior state estimate \( \left( p_k^- \right) \) as well as the covariance of state estimate \( \left( \xi_k^i \right) \) in the prediction step at the edge–cloud are:

\[
p_k^- = F_pp_{k-1} \quad \xi_k^- = F_p\xi_{k-1}F_p^T + Q_p. \] (13a, 13b)

Here, the state transition matrix \( F_p \in \mathbb{R}^{4 \times 4} \) and state noise covariance matrix \( Q_p \in \mathbb{R}^{4 \times 4} \) are similar to those in (8) except that \( \Delta \tau \) now depends on how often positioning is done at the edge cloud, and \( \sigma^2_w \in \mathbb{R} \) is again a design parameter. The posterior covariance estimate \( \xi_k \) and state estimate \( \left( p_k \right) \) in the global EKF engine is:

\[
\xi_k = \left( (\xi_k^-)^{-1} + J(p_k^-) \right)^{-1} \] (14a)

\[
p_k = p_k^- + \xi_kq(p_k^-). \] (14b)

with FIM \( (J(p_k)) \) and score–function \( (q(p_k)) \) given as:

\[
J(p_k) = \frac{2}{\sigma^2_w} \left( \frac{\partial y^p_k}{\partial p_k} \right)^\dagger \left( P_k \right)^{-1} \frac{\partial y^p_k}{\partial p_k} \] (15a)

\[
q(p_k) = \frac{2}{\sigma^2_w} \left( \frac{\partial y^p_k}{\partial p_k} \right)^\dagger \left( P_k \right)^{-1} (m_k - y^p_k). \] (15b)

Here, matrix \( P_k \in \mathbb{R}^{2 \times 2} \) denotes the covariance of \( y^p_k \) \in \mathbb{R}^2 \) at time–instant \( k \), which is given as:

\[
P_k = \begin{bmatrix} \sigma^2_k^A & 0 \\ 0 & \sigma^2_k^B \end{bmatrix}.
\] (16)

Here, \( \sigma^2_k^A \in \mathbb{R} \) and \( \sigma^2_k^B \in \mathbb{R} \) denote the estimated variance of the azimuth angle estimates, obtained from the local EKFs at TRPs \( A \) and \( B \), respectively. Moreover, \( m_k = [\phi_k^A, \phi_k^B]^T \) denotes the measurement vector, also obtained from the local EKFs from TRP \( A \) and TRP \( B \). Finally, \( y^p_k \in \mathbb{R}^2 \) denotes the measurement model vector, which is given as:

\[
y^p_k = \begin{bmatrix} \arctan(\frac{y^\text{UE} - y^\text{TRP}_A}{x^\text{UE} - x^\text{TRP}_A}) \\ \arctan(\frac{y^\text{UE} - y^\text{TRP}_B}{x^\text{UE} - x^\text{TRP}_B}) \end{bmatrix}^T.
\] (17)

Here, \((x^\text{TRP}_A, y^\text{TRP}_A)\) and \((x^\text{TRP}_B, y^\text{TRP}_B)\) denote the known locations of TRP \( A \) and TRP \( B \), respectively.

C. THEORETICAL PERFORMANCE BOUND ON ANGLE ESTIMATION

It is important to know the limit on the accuracy of AoA estimates. The accuracy of an estimator can be described by comparing the estimate with ground truth or using the variance of the estimates. Given the data model described in Section II, the general deterministic Cramér–Rao lower bound (CRLB) on the covariance matrix of any unbiased estimator of \( \theta \) is given as [27, 28]:

\[
\text{CRLB}_{\theta} = \frac{\sigma^2_w}{2} \left| \mathbb{E} \left[ S^H D \Pi_A^T D S \right] \right|^{-1}
\] (18)

Here, \( \Pi_A^T \) denotes the projection onto nullspace of \( A \) read as \( \Pi_A^T = \mathbb{I} - A(A^HA)^{-1}A^T \) where \( A \) is the beampattern of \( m \) far–field narrow–band sources, and \( D \) is the partial derivative of \( A \) w.r.t. \( \theta \), both given as:

\[
A = [B_p(\phi_1, \tau_1), B_p(\phi_2, \tau_2), \ldots, B_p(\phi_m, \tau_m)] \] (19a)

\[
D = [B'_{p_1}, B'_{p_2}, \ldots, B'_{p_m}], B'_{p_m} = \frac{\partial B_{p_m}}{\partial \phi_m}. \] (19b)

Although the expression for the CRLB in (18) is explicit, it is somewhat cumbersome. More insight can be obtained by considering the utilized ULA with single user. The asymptotic CRLB on the variance of AoA estimation for two–dimensional single–path case having independent and identically distributed (i.i.d.) circular Gaussian noise, exponential model, and single path weight (with real and imaginary components) is given as [25]:

\[
\text{CRLB}_{\phi} = \frac{\sigma^2_w}{|\xi|^2} \sqrt{\frac{6\lambda^2}{(2\pi d^2)^2 \sin^2(\phi\delta T) M_T N^2 - 1}},
\] (20)

where \( M_T \) denotes the total number of samples, given as \( M_T = N\delta T \) for one TTI realization. It is clear to see from (20) that as the signal–to–noise ratio (SNR) term \( \frac{\sigma^2_w}{\sigma^2_c} \) increases, the CRLB reduces. Besides, as the array size increases, one can form a better estimate. Furthermore, the more samples \( M_T \) we have, the better the estimate we can obtain. Finally, the \( \sin(\phi) \) term represents the fact that as one scans off broadside, the beamwidth increases, i.e., beam broadening factor making AoA estimates much worse.
FIGURE 2. Our testbed and measurement environment.

IV. TESTBED DESCRIPTION AND EXPERIMENTAL SETUP
A. DESCRIPTION OF UDN TESTBED
Our testbed comprises two TRPs, each composed of a standard host computer and two Universal Software Radio Peripherals (USRPs) X300 series with UBX–160 and SBX RF–daughterboards as the radio front–end. This is shown in Fig. 2 where ULAs of four elements are situated on lamp posts, with inter–element spacing (λ). The array response of each antenna element was measured in an anechoic chamber for all azimuth and elevation angles and the EADF of these practical antenna array was constructed from these measurements. The USRPs perform conversion between baseband and RF signals operating at a center frequency of 3.42 GHz. A license for using this frequency in our campus area was obtained. The USRPs are fed with a common external 10 MHz reference clock and pulse per second (PPS) trigger. A separate transmitter, transmitting known signal to each TRP, is also used as a reference to measure the phase offset due to the clock drift. The TRPs perform run time compensation by computing the phase difference between RF chains, based on the received reference signal from this separate transmitter.

The baseband inphase/quadrature (I/Q) samples are transported to and from the host computer over a 10G Ethernet link. All baseband processing, a subset of physical layer protocol implementation including initial synchronization of UE, periodical tracking of UE’s synchronization w.r.t TRP for sample and frame level alignment for TDD system, PSP transmission from UE, and AoA estimation at TRP run as software on an Ubuntu Linux–based desktop computer. Further detail on the main software components of TRP can be found in [29]. Utilized testbed’s air interface parameters and setup information are summarized in Table 1.

B. EXPERIMENTAL SETUP
The baseline layout used to validate our work is shown in Fig. 3. In this setup, TRPs are placed ISD meters apart from each other on the lamp posts in a parking lot, and a mobile UE is situated Y meters away from the axis joining both TRPs. Our goal was to identify the range of azimuth angles that allow us to achieve sub–meter positioning accuracy. In particular, the following cases were considered: 1) \( ISD = 1.5 \times Y = 21.3 \text{ m, } Y = 14.20 \text{ m} \) 2) \( ISD = 1.2 \times Y = 17.5 \text{ m, } Y = 14.58 \text{ m} \) and 3) \( ISD = Y = 12.5 \text{ m} \). Each

Algorithm 1: channel parameter estimation and tracking
1 Initialization of parameters
2 \( \psi_{grid} \leftarrow [0, \pi] \quad \triangleright \text{initialize angle grid table} \)
3 \( \tau_{grid} \leftarrow [0, 1/f_s] \quad \triangleright \text{initialize time delay grid table} \)
4 \( G \leftarrow \text{EADF} \quad \triangleright \text{load EADF and assign it to } G \)
5 run TRPs \( \triangleright \text{TRPs starts transmitting PSS} \)
6 run UE \( \triangleright \text{UE decodes PSS for synchronization} \)
7 if UE synchronized then
8 \( \hat{h} \leftarrow \text{PSP} \quad \triangleright \text{TRPs perform channel estimate from PSSs as in (2)} \)
9 if TTI number == 1 then
10 11 \( \triangleright \text{start of initial estimation of } \tau \text{ and } \varphi \)
12 \( H \leftarrow \text{reshape}(\hat{h}, N_f, N) \quad \triangleright \text{reshape } \hat{h}^{N_f \times N} \) to \( H^{N_f \times N} \)
13 \( a_v \leftarrow H * (G_v)^\dagger \) and \( a_h \leftarrow H * (G_h)^\dagger \quad \triangleright \text{correlate } H \text{ with } G_v \text{ and } G_h \)
14 \( A_v \leftarrow \text{reshape}(a_v, N_f, M_d, M_c) \)
15 \( A_h \leftarrow \text{reshape}(a_h, N_f, M_d, M_c) \quad \triangleright \text{reshape } a_v \text{ and } a_h \)
16 where \( M_d \text{ and } M_c \text{ are the number of modes} \)
17 \( B_v \leftarrow |\text{FFT}_2D(A_v)|^2 \quad B_h \leftarrow |\text{FFT}_2D(A_h)|^2 \quad \triangleright \text{employ 2D FFT on } A_v \text{ and } A_h \)
18 and obtain \( B \text{ from } B_v \text{ and } B_h \)
19 \( (i^*, j^*) \leftarrow \text{argmax } B(i, j) \quad \triangleright \text{find indices of matrix } B \text{ that result in maximum value} \)
20 \( \tau \leftarrow \tau_{grid}(i^*) \quad \varphi \leftarrow \varphi_{grid}(j^*) \quad \triangleright \text{the delay and AoA grid values corresponding to indices } i^*, \text{ and } j^* \text{ the channel parameter estimates.} \)
21 \( \triangleright \text{end of initial estimate of } \tau \text{ and } \varphi \)
22 else
23 \( \triangleright \text{start of EKF–based tracking of } \tau \text{ and } \varphi \)
24 Obtain \( \theta_k^- \) using (7) \( \triangleright \text{prediction step} \)
25 Obtain \( \theta_k^+ \) using (9) \( \triangleright \text{update step based on } \theta_k^- \text{ and measurement} \)
26 \( \tau \leftarrow \theta_k^+(1) \quad \varphi \leftarrow \theta_k^+(2) \quad \triangleright \text{the delay and AoA estimates} \)
27 \( \triangleright \text{end of EKF–based tracking of } \tau \text{ and } \varphi \)
28 if TTI number == 2 then
29 \( \Delta \varphi_2 \leftarrow \frac{\varphi_2 - \varphi_1}{\Delta \varphi} \quad \Delta \varphi_2 \leftarrow \frac{\varphi_2 - \varphi_1}{\Delta \varphi} \quad \triangleright \text{Initial rate–of–change parameters} \)
30 \( (\xi_2)_{13} \leftarrow \frac{(\xi_2)_{13} + (\xi_2)_{14}}{\Delta \tau} \quad (\xi_2)_{14} \leftarrow \frac{(\xi_2)_{13} + (\xi_2)_{14}}{\Delta \tau} \quad \triangleright \text{update of covariance matrix} \)
31 end
32 else
33 \( \triangleright \text{UE keeps on trying synchronization and no PSP transmission from UE} \)
34 go back to step 6
of these cases have changing 2D geometry and the incident angles as well. In all cases, the UE movement was taken in discrete steps, from point 1 to point \( K \), or vice-versa, as shown in Fig. 3, where each point is 2.5 m away from its neighbour point. The general movement model described in (7) is valid for continuous movements. In our measurement campaign, the UE was placed in each point and held fixed until data is collected for 1000 TTIs. Then, the UE was moved to another location. The EKF was re-initialized every time the UE was re-located, and tracking was performed for each location of the UE in an independent manner. For the sake of generality, the EKF does not assume the UE to be static, and therefore the rate of change of \((\varphi, \tau)\) are also tracked.

V. MEASUREMENT RESULTS AND DISCUSSION

Performance results obtained from testbed measurements are reported in this section. As described in (20), the performance of an AoA estimator is dependent on SNR value. The average measured SNR result per point, for all measurement cases and TRPs, is as shown in Fig. 4. It is clear to see that the average SNR value in each measurement point on both TRPs is above 20 dB, which is good enough for AoA estimation. Moreover, the corresponding CRLB of \( \varphi \) is computed using (20) and compared with the measured sample variance of AoA estimation as shown in Fig. 5, for the second TRP as an illustrative example.

The first set of measurements was carried out for the case of \( ISD = Y = 12.5 \) m, i.e. the maximum azimuth AoA, when measured from boresight, was 45º. Fig. 6 shows the azimuth AoA estimate vs ground truth at each TRP using the MLE. The MLE provides reasonable results only at certain TTIs. Significant errors are observed for some TTIs due to variation in clock between the UE and TRP as a result of imperfections of the oscillators driving clocks. Although the UE tracks its synchronization periodically every given
Fig. 7 shows the same case as Fig. 6, but by using the EKF–based approach instead of the MLE for estimation and tracking of azimuth AoA. It can be seen that the momentary outliers have been rejected effectively by filtering. Additionally, the computational load while utilizing EKF is small compared to exhaustive search–based MLE approach. Therefore the EKF–based results show that significant improvement can be obtained in terms of performance for latency–constrained services of 5G. Fig. 8 shows the corresponding position estimate at the edge cloud using EKF–based AoA from TRPs for the case of \( ISD = Y \).

The second set of measurements was carried out for the case of \( ISD = 1.23 \ast Y = 17.5 \, m \). Fig. 9 and 10 show the position estimates using MLE and EKF–based approaches respectively for the said case. It is clear to see from Fig. 9 that the position estimate is highly influenced by the difference in the clocks running UE and TRP. On the other hand,
the position estimate in Fig. 10 has better accuracy around the boresight of the array as a result of filtering and the accuracy decreases as one scans off broadside due to the beam broadening effect. The third set of measurements was carried out for the case of \( ISD = 1.5 \times Y = 21.3 \) m, i.e., the azimuth AoA ranges from 0° to 56°. Fig. 11 show AoA estimates at each TRP using the EKF–based approach for the case of \( ISD = 1.23 \times Y = 17.5 \) m. Similarly, the results show that AoA estimation is poor at the end–fire of the array for both TRPs, i.e. when the azimuth AoA is below 45°.

Fig. 12 shows the cumulative distribution function (CDF) of the positioning error for all three cases described earlier. These results are helpful in identifying the usable range of azimuth angles (i.e., \( ISD \) vs \( Y \) relation) of the described testbed scenario for a pre–defined target accuracy. This \( ISD \) vs \( Y \) relation shows the beamwidth. The result shows that sub–meter positioning accuracy can be achieved in about 95% of the considered UE locations, for the \( ISD = Y \) case. Therefore, sub–meter positioning accuracy can be achieved by tackling the effect of beam broadening in the network design. Effectively, the UDN network has to be designed in such a way that the UE’s trajectory (e.g. road users on highways) should fall in the usable azimuth angle range of serving TRPs.

VI. CONCLUSION

This work has addressed experimentally achievable positioning accuracy of connected devices in a 5G UDN using proof–of–concept user–centric positioning testbed. The considered scheme consists of spatial parameters’ estimation at multiple TRPs and subsequent position estimation at an edge–cloud using the well–known EKF–based approach. This work also contributes to the body of knowledge by addressing practical impairments faced in implementation. Specifically, the practical limitations include non–uniformity in radiation pattern among the antenna elements in the array and non–identical characteristics of RF front–ends, antenna feed cables, connector and other RF components. All these impairments have significant effect on the relative phase of signals among received streams, which is the basis for AoA–based positioning approach. To tackle these problems, three step calibration techniques have been implemented in this paper. The first step is aimed at solving time–varying phase–offset problem due to clock drifts. A separate transmitter sends reference signals and then TRPs perform run time compensation by computing the phase difference between RF chains. Second step mitigates the phase offset as a result of non–identical antenna feed cables, connectors and other RF components. This is achieved by performing phase measurement of each component via VNA and accounting for the offset. Third step is aimed at taking into account individual practical antenna element gain and phase characteristics, effect of mutual coupling, cross–polarization and mounting platform reflections. This is achieved by augmenting the theoretical beampattern with EADF which is constructed from the measured antenna response in anechoic chamber.

The results obtained from extensive measurements with our testbed demonstrate that sub–meter 2D positioning accuracy of devices can be achieved with high probability (95%) in realistic UDN scenarios by employing ULAs with at least four antennas per TRP and properly designing the UDN.
Future work will focus on extending the air-interface to a 5G NR standard compliant one and also incorporating data obtained from multiple sensor measurements in order to further improve accuracy of position information.

REFERENCES


RIKU JÄNTTI (M’02–SM’07) received the M.Sc. degree (Hons.) in electrical engineering and the D.Sc. degree (Hons.) in automation and systems technology from the Helsinki University of Technology (TKK), in 1997 and 2001, respectively. He was a Professor pro term with the Department of Computer Science, University of Vaasa. In 2006, he joined the School of Electrical Engineering, Aalto University (formerly known as TKK), Finland, where he is currently an Associate Professor (tenured) in communications engineering and the Head of the Department of Communications and Networking. His research interests include radio resource control and optimization for machine type communications, cloud-based radio access networks, spectrum and co-existence management, and RF inference. He is an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is also an IEEE VTS Distinguished Lecturer (Class 2016).

KALLE RUTTI received the Diploma degree in engineering from Tallinn Technical University, in 1993, and the Lic.Tech. degree from the Helsinki University of Technology, in 1999. He is currently a Teaching Researcher with the Department of Communications and Networking (ComNet), School of Electrical Engineering, Aalto University. His research interests include radio physical layer algorithms, software radio, base station implementation, the NB-IoT, and cloud RAN.

MÁRIO COSTA (S’08–M’13) received the M.Sc. degree (Hons.) in communications engineering from the Universidade do Minho, Portugal, in 2008, and the D.Sc. (Tech.) degree in electrical engineering from Aalto University, Finland, in 2013. In 2014, he was a Visiting Postdoctoral Research Associate with Princeton University. Since 2014, he has been with Huawei Technologies Oy (Finland) Co., Ltd., as a Senior Researcher. His research interests include statistical signal processing and wireless communications.

KARI LEPPÄNEN received the M.Sc. and Ph.D. degrees from the Helsinki University of Technology, Finland, in 1992 and 1995, respectively, majoring in space technology and radio engineering. He was with National Radio Astronomy Observatory, USA, Helsinki University of Technology, Finland, Joint Institute for VLBI in Europe, The Netherlands, Nokia Research Center, Finland, and Huawei Finland. His research interests include radio algorithms, and protocols for ultra-dense wireless networks and radio positioning.

* * *