Abstract—Emerging cellular networks integrate the user terminal geo-localization function besides the communication function. The conventional positioning approach is to estimate the terminal location in two-steps: first the distance to all connected base stations is assessed based on signal time-of-flight measurements, then the location is deduced from the distances by multi-lateration. The two-step approach incurs a performance degradation because information is lost from the received signal when the multi-lateration is performed. In this paper, we propose to iterate between the two conventional steps to progressively refine the distance estimates based on the knowledge of the position estimate obtained from the previous iterations. The information exchanged between the two-steps not only consists in the mean of the estimates (distance or position) but also of their variance that convey information about the reliability of the estimates. Simulation results show that the achievable performance after a few iterations is close to the performance of the optimal approach that directly estimates the position based on the observation of the received signal.

Index Terms—Localization, iterative processing

I. INTRODUCTION

From the nineties onward, cellular communications networks have continuously evolved to become one of the main blocks of our ICT all-pervasive world. In parallel, they have also evolved towards geo-located services [1]. In 2G and 3G, the localization process is very crude, and mainly limited to cell ID localization typically used to accelerate the initialization phase of the more accurate global navigation satellite system (GNSS) positioning. In 4G, localization has become an essential part of cellular network functionality that may even replace the need for the additional GNSS. A positioning reference signal (PRS) is included in the protocol to support the device localization based on the estimation of the signal time-of-flight (ToF) to the base stations. The PRS is defined as an orthogonal frequency-division multiplexing (OFDM) signal scattered in time and frequency [2].

Time-of-arrival (ToA) algorithms employ the information of the absolute signal ToF from the transmitting terminal to the receiving base stations (or reciprocally). This approach requires the knowledge of signal departure time and thus synchronization between the transmitter and receivers [3]. The unknown time of emission may alternatively be estimated together with the transmitter position [4]. Time Difference of Arrival (TDoA) is preferred if there is no synchronization between the transmitter and the receivers. In this case, only the relative signal travel times are known and the difference of ToF between the links is used instead of the absolute ToF. The conventional localization approach proceeds in two steps: (i) the ToA/TDoA estimation based on the received signals; (ii) the position estimation obtained by multi-lateration based on the ToA/TDoA. The time estimates obtained at each receiver at the output of the first step are communicated to a common processing center where the device position is estimated in the second step. Finding the location based on the estimated ToA/TDoA is not trivial because these measurements have non-linear relationships with the source position. A lot of effort has been devoted to design efficient algorithms to solve the system of non-linear equations. They either work directly with the non-linear equations or transform the relationships to obtain an approximated set of linear equations. The least square (LS) and maximum likelihood (ML) estimators are detailed for the two cases in [5].

The two-step localization approach is suboptimal since information is lost from the received signals when the ToA/TDoA estimates are communicated to the multi-lateration step (i.e. ToA/TDoA is not a sufficient statistic). Paper [6] demonstrates that the positioning accuracy can significantly be improved by directly estimating the location based on the received signals. The ML estimator is derived for the cases of known and unknown transmitted signals. The results are extended in [7] to further account for the unknown transmission time. Interestingly, [8] analytically demonstrates that the direct estimation approach always outperforms the two-step approach. Simulations results show a significant performance gain, especially at low signal-to-noise power ratios (SNR) [9].

While the direct position estimation brings interesting performance gains, the method unfortunately suffers from a significant computational complexity increase. Also the full knowledge of the received signals is necessary at the central processing center, incurring the need for high capacity...
control channels. In this paper, we propose to still perform the localization in two distinct steps but to iterate between them to progressively refine the estimates. The iterative ToA-based positioning makes use of the Bayes framework to take benefit at each iteration from the prior knowledge obtained from the former iteration. The information exchanged between the two steps is composed of the two first order moments of the estimates (mean and variance of the time or position estimates). Although the concept can be applied in most of the localization scenarios, we focus on emerging cellular systems based on OFDM-like modulations and composed of small cells. The main goal is to show the feasibility of the approach and to assess the performance compared to the conventional two-step and direct positioning approaches.

The rest of the paper is organized as follows. Section II first introduces the OFDM signal model. Section III secondly details the iterative positioning algorithm. After an overview of the algorithm, each step is detailed mathematically. Section IV finally assesses the performance of the iterative algorithm numerically and compares it to the state-of-the-art approaches. In the text, vectors and matrices are identified by using lowercase and uppercase bold letters respectively.

II. SIGNAL MODEL

We consider the uplink of a cellular communication system composed of small cells. The results are straightforwardly extended to the downlink. A user terminal is simultaneously connected to $K$ different base stations in its neighbourhood. The communication takes place in a bandwidth $B$ centered around the carrier frequency $f_c$. The OFDM modulation, used in 4G cellular communication systems, is assumed. The principle of OFDM is to sub-divide the frequency spectrum in $Q$ sub-carriers allocated to either data or pilot symbols. A cyclic prefix (CP) is added to each block of transmitted symbols to maintain the orthogonality between the sub-carriers even in the presence of channel time dispersion.

Assuming for simplicity that the channel is a single propagation delay $\tau_k$ between the user terminal and the base station $k$, the baseband channel response is a delayed dirac pulse times a constant phase rotation due to the carrier frequency shift, as expressed by $\delta(t - \tau_k) e^{-j2\pi f_c t \tau_k}$. The delay is equal to $\tau_k = c d_k$ where $c$ is the speed of light and $d_k$ is the distance separating the user terminal from the base station $k$. If $\tau_k$ is shorter than the CP duration, the signal received on the sub-carrier $q$ at the base station $k$ is:

$$r_{kq} = s_q e^{-j2\pi \frac{q \tau_k}{Q}} e^{-j2\pi f_c \tau_k} + w_{kq}$$  \hspace{1cm} (1)

for $k = 1 \cdots K$ and $q = -Q/2 \cdots Q/2 - 1$. In (1), $s_q$ is the data/pilot symbol transmitted on the sub-carrier $q$, $T = 1/B$ is the sample duration and $w_{kq}$ is the noise corrupting the sub-carrier $q$ at the base station $k$. Additive white Gaussian noise (AWGN) of variance $\sigma_w^2$ is assumed.

The propagation delay causes a constant phase rotation over the sub-carriers and a phase rotation proportionally increasing with the sub-carrier index. We neglect the constant phase rotation later on as it is usually pre-compensated together with the phase difference existing between the transmit/receive local oscillators before the positioning takes place. The delay is therefore estimated by observing the phase slope on a subset $\mathcal{P} = \{q_1 \cdots q_P\}$ of received pilot sub-carriers. Because the OFDM subcarriers are orthogonal, no interference is caused by the information data on the pilot signals. An equivalent vector model can be built by gathering the observations on the pilot carriers in a vector:

$$r_k = s(\tau_k) + w_k$$ \hspace{1cm} (2)

where

$$r_k := \begin{bmatrix} r_{kq_1} & \cdots & r_{kq_P} \end{bmatrix}^T$$ \hspace{1cm} (3)

$$w_k := \begin{bmatrix} w_{kq_1} & \cdots & w_{kq_P} \end{bmatrix}^T$$ \hspace{1cm} (4)

and

$$s(\tau_k) := \begin{bmatrix} s_{q_1} e^{-j2\pi q_1 \frac{\tau_k}{Q}} & \cdots & s_{q_P} e^{-j2\pi q_P \frac{\tau_k}{Q}} \end{bmatrix}^T.$$ \hspace{1cm} (5)

III. ITERATIVE POSITIONING

A. Algorithm description

Fig. 1 illustrates the functional blocks involved in the iterative delay/position estimation and the information exchanged at each iteration between the blocks. The two main blocks are the delay estimation and the position estimation. The other blocks are necessary to translate the output of each of the two main blocks to information relevant for the other block.

The 'delay estimation' functional block is independently implemented at each base station and aims at estimating the signal propagation delay from the user terminal to the base station. It refines prior information received on the delay (the delay mean and variance) based on the observation of the OFDM received signal on the pilot sub-carriers. As the number of iterations increases, the additional information brought by the observation of the received signal becomes less

![Fig. 1. Iterative delay/position estimation.](image-url)
relevant compared to the more reliable prior information. The delay estimation algorithm is detailed in Section III-B. The information on the delays is easily converted to information on the distances to the base station by multiplication with the speed of light.

One instance of the ‘position estimation’ functional block is implemented per base station at the common signal processing center. It produces an estimate of the user terminal position expressed in the \((x, y)\) plane (the means) along with its reliability (the variances) based on the information on the distances provided by all other base stations. This ensures the independence of the estimate with respect to the information coming from the base station itself at the last iteration. The position estimation algorithm is detailed in Section III-C. The information on the user terminal position must be converted in information on the signal propagation delay to the base station, which is performed successively in two steps: position to distance, distance to delay. The relationship between the position and the distance is non-linear, making the translation more difficult. A linear approximation is proposed in Section III-D. The information on the distances is easily converted to information on the signal propagation delays by division with the speed of light.

B. Delay estimation step

The delay \(\tau_k\) is estimated at each base station separately based on the observation of the received vector (2). Making use of the Bayes framework [10], we assume that \(\tau_k\) is a random variable characterised by a prior probability density function (PDF). If some prior knowledge about the position of the user terminal is available, it is realistic to assume that \(\tau_k\) follows a Gaussian distribution completely specified by its first two moments: the mean \(\mu_{\tau_k}\) and the variance \(\sigma_{\tau_k}^2\).

The posterior distribution is therefore:

\[
p(\tau_k | r_k) = \frac{p(\tau_k | r_k)p(\tau_k)}{\int_{-\infty}^{\infty} p(\tau_k | r_k)p(\tau_k) d\tau_k} = \frac{\exp \left( \frac{1}{\sigma_{\nu}^2} \Re \left( r_k^H \cdot s(\tau_k) \right) + \frac{1}{\sigma_{\tau_k}^2} \left( \mu_{\tau_k} - \frac{\tau_k - c}{2} \right) \right)}{\int_{-\infty}^{\infty} \exp \left( \frac{1}{\sigma_{\nu}^2} \Re \left( r_k^H \cdot s(\tau_k) \right) + \frac{1}{\sigma_{\tau_k}^2} \left( \mu_{\tau_k} - \frac{\tau_k - c}{2} \right) \right) d\tau_k}
\]

where \(\Re(\cdot)\) denotes the real part operator. The expression (7) has been obtained by inserting the expressions of the Gaussian distributions \(p(r_k | \tau_k)\) and \(p(\tau_k)\) and by simplifying the terms independent of \(\tau_k\) as they appear both at the numerator and denominator. The denominator makes sure the PDF integrates to one. It is interesting to note that the variances \(\sigma_{\nu}^2\) and \(\sigma_{\tau_k}^2\) weight the two terms at the numerator. When the SNR is small (\(\sigma_{\nu}^2\) is large), more importance is given to the prior information in the estimation of \(\tau_k\) (second term). Conversely more importance is given to the information gained by the observation of the received vector (first term) when the SNR is large (\(\sigma_{\tau_k}^2\) is small).

It is possible to numerically compute the mean and variance of the variable \(\tau_k\) based on the knowledge of the posterior distribution:

\[
\mu_{\tau_k | r_k} = \mathbb{E} [\tau_k | r_k] = \int_{-\infty}^{\infty} \tau_k p(\tau_k | r_k) d\tau_k \quad (8)
\]

and

\[
\sigma_{\tau_k | r_k}^2 = \mathbb{E} \left[ (\tau_k - \mu_{\tau_k | r_k})^2 | r_k \right] = \int_{-\infty}^{\infty} (\tau_k - \mu_{\tau_k | r_k})^2 p(\tau_k | r_k) d\tau_k \quad (9)
\]

where \(\mathbb{E}[\cdot]\) denotes the statistical expectation operator. Knowing \(\mu_{\tau_k | r_k}\) and \(\sigma_{\tau_k | r_k}^2\) the mean and variance of the corresponding distance \(d_k\) between the user terminal and base station \(k\) are easily deduced:

\[
\mu_{d_k} = c \mu_{\tau_k | r_k} \quad (11)
\]

\[
\sigma_{d_k}^2 = c^2 \sigma_{\tau_k | r_k}^2 \quad (12)
\]

In the position estimation step, the mean is used as observed distance (minimum mean square error (MMSE) estimate [10]) and the variance is used as the variance of the error on the observation.

C. Position estimation step

As explained in Section III-A, one position estimate in the \((x, y)\) plane is computed per base station based on the distances observed by the other \(K-1\) base stations. This ensures the independence of the prior information communicated to the delay estimation step in the next iteration with the signal received at the base station.

If the base station \(k\) is located at the position \((x_k, y_k)\), its observed distance \(d_k\) is the actual distance \(d_k\) corrupted by an additive error \(e_k\) also assumed to be Gaussian of mean equal to 0 and variance equal to \(\sigma_{e_k}^2\):

\[
\hat{d}_k = d_k(x, y) + e_k \quad (13)
\]

Similarly to what has been done before to estimate the delays, the observations at the \(K-1\) base stations can be gathered to form a vector model (for the sake of clarity, we do not emphasize the fact that the index corresponding to the target base station should be absent in the vector expressions):

\[
\hat{d} = d(x, y) + e \quad (14)
\]

where

\[
\hat{d} := \begin{bmatrix} \hat{d}_1 & \cdots & \hat{d}_K \end{bmatrix}^T \quad (15)
\]

\[
d(x, y) := \begin{bmatrix} d_1(x, y) & \cdots & d_K(x, y) \end{bmatrix}^T \quad (16)
\]

\[
e := \begin{bmatrix} e_1 & \cdots & e_K \end{bmatrix}^T \quad (17)
\]

The elements of \(e\) are assumed to be independent but not necessarily of the same variance to take possible different distance estimation reliabilities at the different base stations into account. The covariance matrix is diagonal and written as (same remark regarding the index of the target base station):

\[
\mathbf{R}_e = \mathbb{E} [e \cdot e^H] = \begin{bmatrix} \sigma_{e_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{e_K}^2 \end{bmatrix} \quad (18)
\]
We also make use of the Bayes framework to estimate the position assuming that the user terminal can arbitrarily be located on a finite rectangle in the plane. Therefore, \( x \) is a uniform random variable on \([X_{\min}, X_{\max}]\) and \( y \) is a uniform random variable on \([Y_{\min}, Y_{\max}]\).

The posterior PDF is given by:

\[
p(x, y | \hat{d}) = \frac{p(\hat{d}|x, y)p(x, y)}{\int_{X_{\min}}^{X_{\max}} \int_{Y_{\min}}^{Y_{\max}} p(\hat{d}|x, y)p(x, y)dx dy}
\]

\[
= \frac{\exp \left( \frac{1}{2}(\hat{d} - \frac{1}{2}d(x, y))^T \cdot R^{-1} \cdot d(x, y) \right)}{\int_{X_{\min}}^{X_{\max}} \int_{Y_{\min}}^{Y_{\max}} \exp \left( \frac{1}{2}(\hat{d} - \frac{1}{2}d(x, y))^T \cdot R^{-1} \cdot d(x, y) \right) dx dy}
\]

for \( x \in [X_{\min}, X_{\max}] \), \( y \in [Y_{\min}, Y_{\max}] \) and 0 elsewhere.

It is possible to numerically compute the mean and (co)variance of the variables \( x, y \) based on the knowledge of the posterior distribution:

\[
\mu_{x|\hat{d}} = E \left[ x | \hat{d} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x p(x, y | \hat{d}) dx dy
\]

\[
\mu_{y|\hat{d}} = E \left[ y | \hat{d} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y p(x, y | \hat{d}) dx dy
\]

and

\[
\sigma_{x|\hat{d}}^2 = E \left[ (x - \mu_{x|\hat{d}})^2 | \hat{d} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_{x|\hat{d}})^2 p(x, y | \hat{d}) dx dy
\]

\[
\sigma_{y|\hat{d}}^2 = E \left[ (y - \mu_{y|\hat{d}})^2 | \hat{d} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_{y|\hat{d}})^2 p(x, y | \hat{d}) dx dy
\]

\[
\Gamma_{xy|\hat{d}} = E \left[ (x - \mu_{x|\hat{d}})(y - \mu_{y|\hat{d}}) | \hat{d} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_{x|\hat{d}})(y - \mu_{y|\hat{d}}) p(x, y | \hat{d}) dx dy
\]

The point \((\mu_{x|\hat{d}}, \mu_{y|\hat{d}})\) in the plane constitutes a relevant estimate of the user terminal position (MMSE estimate [10]).

D. Position to distance translation

Knowing the mean \((\mu_x, \mu_y)\) and (co)variance \((\sigma_x^2, \sigma_y^2, \Gamma_{xy})\) of the coordinates \((x, y)\), the first two order moments of the distance \(d_k\) to each base station \(k\) can be computed. Unfortunately, the relationships between the coordinates and the distances are non-linear making the computation less straightforward. A linear approximation around the mean is performed in order to obtain closed-form expressions. Assuming a small variation of the coordinates around their mean, which makes sense since the variances \(\sigma_x^2\) and \(\sigma_y^2\) are generally small, we have that:

\[
x = \mu_x + \Delta x
\]

\[
y = \mu_y + \Delta y,
\]

and it is easily shown that:

\[
d_k(x, y) \approx d_k(\mu_x, \mu_y) - \frac{x_k - \mu_x}{d_k(\mu_x, \mu_y)} \Delta x - \frac{y_k - \mu_y}{d_k(\mu_x, \mu_y)} \Delta y.
\]

Therefore the mean of \(d_k(x, y)\) is approximately:

\[
\mu_{d_k} = d_k(\mu_x, \mu_y)
\]

and the variance of \(d_k(x, y)\) is approximately:

\[
\sigma_{d_k}^2 \approx \frac{1}{d_k^2(\mu_x, \mu_y)} \left[ \frac{\sigma_x^2}{\Gamma_{xy}} \right]^T \left[ \begin{array}{cc} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{array} \right] \left[ \begin{array}{c} x_k - \mu_x \\ y_k - \mu_y \end{array} \right]
\]

because \(E[\Delta x] = E[\Delta y] = 0, E[\Delta x^2] = \sigma_x^2, E[\Delta y^2] = \sigma_y^2\) and \(E[\Delta x \Delta y] = \Gamma_{xy}\).

Knowing \(\mu_{d_k}\) and \(\sigma_{d_k}^2\), the mean and variance of the corresponding delay \(\tau_k\) between the user terminal and base station \(k\) are easily deduced:

\[
\mu_{\tau_k} = \frac{1}{c} \mu_{d_k}
\]

\[
\sigma_{\tau_k}^2 = \frac{1}{c^2} \sigma_{d_k}^2.
\]

The two estimated moments are used as prior information in the delay estimation step.

IV. NUMERICAL RESULTS

A user terminal arbitrarily lies in a square of side length equal to 100 meters. A base station is located at the four corners of the square. The user terminal communicates in the uplink with the four base stations in a 40 MHz bandwidth using the OFDM modulation. The number of OFDM subcarriers is equal to 1024. There is a pilot symbol transmitted for localization purposes every 16 sub-carriers. The SNR is assumed to be equal at the four base stations. The performance is averaged over 1000 terminal position and noise realizations.

Fig. 2 illustrates the average localization error as a function of the SNR. The iterative position estimation is compared to the conventional two-step estimation and to the direct estimation. At the first iteration, the iterative estimation performs slightly worse than the conventional two-step estimation. In case of the iterative estimation, the position is estimated by...
averaging the estimate obtained with the four possible sets of 3 base stations, while the 4 base stations are directly used in case of the two-step estimation. However, when the number of iterations increases, the performance of the iterative estimation comes very close to the performance of the direct position estimation.

The cumulative distribution function (CDF) of the localization error is given in Fig. 3 for a SNR fixed to −12 dB. The conclusions obtained based on Fig. 2 are confirmed: while the performance of the iterative estimation is close to the performance of the conventional two-step estimation at the first iteration, it comes close to the performance of the direct estimation when the number of iterations increases to 10.

Fig. 4 finally illustrates the average localization error as a function of the SNR.