A Propagation-Centric Transmitter Localization Method for Deriving the Spatial Structure of Opportunistic Wireless Networks

Andreas Achtzehn, Ljiljana Simić, Peter Gronerth and Petri Mähönen
Institute for Networked Systems
RWTH Aachen University
Kackertstrasse 9, D-52072 Aachen, Germany
Email: {aac, lsi, pgr, pma}@inets.rwth-aachen.de

***Abstract*—The design of emerging multi-tier dense wireless networks, which integrate opportunistically deployed devices into legacy infrastructure networks, will hugely benefit from a thorough understanding of the spatial structure of existing large-scale unplanned deployments. However, detailed and precise datasets which would enable acquiring such knowledge are unavailable and non-trivial to generate. In this paper we develop a new localization method that focuses on the derivation of transmitter distributions over large areas, such as those of opportunistic Wi-Fi networks. While prior work has emphasized single-node localization through geometrical inference on points of visibility and involved only rudimentary radio propagation modelling, we combine these two approaches in a hybrid technique to improve localization accuracy. Through exploitation of location information for a subset of known transmitters, we derive the parameters of a statistical propagation model that reduces the error induced by shadowing in realistic environments. After deriving theoretical bounds on propagation estimation from measurements, we compare our method to centroid-based localization approaches via simulation. We further assess our algorithm using data from a real Wi-Fi measurement campaign we have carried out in a suburban environment. Our results indicate a significant improvement in localization accuracy using our hybrid approach compared to conventional geometry-focused techniques.**

I. Introduction

The increasing demand for high-speed data communications calls for the emerging paradigm of multi-tier dense wireless networks. As a means of alleviating the imminent wireless capacity problem, mobile network operators are working on data offloading strategies that integrate opportunistically deployed RAN entities such as femtocell access points (APs) into their legacy network infrastructure [1]. In order to optimize radio resource management in such scenarios, a thorough understanding of the underlying spatial structure of randomly deployed wireless networks is invaluable [2], yet currently underdeveloped. Studying existing unplanned deployments, e.g. the distribution of Wi-Fi APs, is key to enabling theoretical and practical assessments of new algorithms, transceiver technologies, and access schemes in cognitive radio and self-organized networks research [3].

In order to generate reliable node location data for deriving the spatial distribution of a large number of transmitters it is necessary to employ precise localization algorithms. A prominent example of this approach is the localization of Wi-Fi APs via large-scale measurement campaigns.

Whereas there is an abundance of literature on the self-localization of devices using signals from known Wi-Fi APs, significantly fewer works are concerned with localization of Wi-Fi APs themselves and the derivation of their distribution. The largest study reported thus far was conducted by Jones et al. [4] and derived AP density statistics for metropolitan areas in the US. However, the geometric structure of observed APs is not reported on. Moreover, although the scale of the study is unmatched, the localization methodology employed by the commercial provider SkyHook Wireless is unfortunately not disclosed. A number of minor studies covering single districts or small cities exist, e.g. [5], which use commercial off-the-shelf (COTS) measurement devices but do not describe in detail the algorithms employed to localize the APs.

Some work has also focused on the spatial statistics of opportunistic networks. In [6], Wi-Fi deployments in Portland, Oregon, were studied to derive connectivity probabilities. However, the focus of this work was on selecting optimum measurement locations rather than accurately localizing the APs themselves. A similar theory-driven study was reported in [7] to derive additional spatial distribution statistics without deriving the actual geometry of AP locations.

Studies in [8]–[11] are more closely related to our work, as they specifically focus on the performance of various localization algorithms. The accuracy of geometric centroid localization algorithms with RSSI weighting is analyzed in [8] for the case of data obtained via driving a car ("wardriving") compared to via walking ("warwalking"). Several modifications to the localization algorithms are discussed, however the analysis is strictly empirical and limited to evaluating the performance using a single dataset without proving their fitness in general. The most closely related work to ours was recently reported in [11]. The authors derive relative AP localizations using dissimilarity between RSSI observed from different APs. However, their algorithm makes the assumptions of transmitter homogenity with equal transmit power and an indoor setting with homogeneous wall shielding. By contrast, our work makes no such restrictive assumptions. Finally, we mention [9], [10] which propose localization algorithms

using data gathered with non-COTS equipment (i.e. directional and steerable beam antennas). However, these studies are concerned with the localization of a single AP only, and it is non-obvious how their algorithms could be extended to the multi-AP case.

In this paper we propose a novel propagation-centric localization algorithm which combines concepts from geometric techniques with an explicit awareness of the radio environment. We show that our hybrid algorithm outperforms conventional centroid techniques, while remaining flexible enough to be applicable in large-scale measurement campaigns using COTS equipment. Moreover, to the best of our knowledge our algorithm is the first technique which integrates measurements from a small set of transmitters with known locations to improve the localization of the remaining transmitters in a large-scale opportunistic network deployment.

The remainder of this paper is organized as follows. In Section II we define our system model and problem statement. In Section III we discuss the shortcomings of conventional geometry-based localization algorithms which only implicitly take into account the radio propagation environment, as a means of motivating the development of our algorithm. In Section IV we present our novel propagation-centric localization algorithm and present a baseline performance analysis. In Sections V and VI we refine our algorithm by separating the localization and radio propagation parameter estimation using a single and multiple known APs, respectively. In Section VII we validate the effectiveness of our proposed localization algorithm using data from a real measurement campaign. Section VIII concludes the paper.

II. System Model

We consider a typical campaign for the localization of multiple transmitters via drive-by measurements with a vehicle-mounted antenna or a pedestrian walking in non-accessible areas. No special assumptions are made about the hardware used in the campaign, i.e. self-localization may be carried out using, e.g. GPS devices or maps, and the signal strength is measured with COTS components. A common scenario of this kind is the localization of IEEE 802.11 Wi-Fi access points (APs), hence we will use the terms *transmitter* and *access point* in the following interchangeably. Our methods are nevertheless applicable to various radio systems.

A set of APs $\mathcal{T}$ is opportunistically distributed over the measurement area $\mathcal{A}$ at locations $H = \{\eta_1, \eta_2, \ldots, \eta_n\} \in \mathcal{A}$, where $\eta_i \in \mathbb{R}^2$ is the location of AP $i \in \mathcal{T}$. Each AP sends a uniquely identifiable signal with power P^i_{TX}. Measurements are carried out at known locations $N = \{\nu_1, \nu_2, \ldots, \nu_m\} \in \mathcal{A}, \nu_i \in \mathbb{R}^2$. The locations of a small subset of APs are known, e.g. from mapping of identifiers to an ISP's customer database. The aim is then to derive the location of all other APs from the measurements.

During the measurement campaign the received signal power of the observable APs is retrieved, represented by a set of tuples $\mathcal{M} = \{(i, \nu_j, P^i_{\text{RX},\nu_j})\}$, where P^i_{RX,ν_j} is the received signal strength of AP i at location ν_j. Signal strength estimation and self-localization of the measurement equipment are carried out carefully, i.e. we assume there is no additional measurement error. The signal strength estimation is constrained by the decodability threshold ψ of the measurement equipment, so that a given transmitter i is "visible" only at a subset of measurement locations N^i, corresponding to the set of tuples $\mathcal{M}^i \subset \mathcal{M}$.

The relation between transmitted and received signal power is modelled using a log-distance pathloss model [12] as

$$P^i_{\text{RX},\nu_j}[dBm] = P^i_{\text{TX}}[dBm] - L \quad (1)$$

$$L[dB] = L_{\text{off},i} + 10\kappa \log_{10} ||\eta_i, \nu_j|| + \chi_{\eta_i,\nu_j}, \quad (2)$$

where $||\cdot,\cdot||$ is the Euclidean distance between locations. $L_{\text{off},i}$ represents the fixed reference attenuation unique to each individual i, incorporating all effects of individual wall shielding, frequency, antenna heights, antenna gains, etc. The pathloss coefficient κ on the other hand is assumed to be the same for all transmitters in $\mathcal{A}$ and represents the extent of additional attenuation given the link distance and propagation environment. Its value is scenario-dependent and usually varies between 1.7 and 4 for rural and urban environments, respectively [12]. On each link, shadowing is modelled through an i.i.d. random variable $\chi_{\eta_i,\nu_j} \sim \mathcal{N}\left(0, \sigma^2\right)$.

Localization Problem Formulation

Using $\mathcal{M}$ and the information from APs with known locations, we aim to find a location estimator $\hat{\eta}_i$, so that the sum square location error $\sum_{i=1}^{n} ||\hat{\eta}_i, \eta_i||^2$ for all APs is minimized.

III. Shortcomings of Radio Environment-Ignorant Geometric Localization Algorithms

Transmitter localization algorithms use the information in $\mathcal{M}$ and apply a prediction model that - implicitly or explicitly - takes physical properties of the propagation environment into account. Simple models take the locations of observation points for each single transmitter, and build a *geometric estimator* for the transmitter location for this data. They thereby exploit the decodability threshold and symmetry of a transmitter's radiation pattern. Signal strength data is generally used for weighting only, and the models do not make use of an underlying pathloss model. By contrast, the method we will propose in this paper takes radio propagation explicitly into account and localizes transmitters by jointly considering the geometry and radio propagation. In this section we discuss the shortcomings of geometric localization techniques that motivate our algorithm.

Localization algorithms take as their input data gathered in a measurement campaign, comprising the tuple set $\mathcal{M}$ as defined in Section II. The signal from AP i is thus visible within the set of points N^i in the vicinity of i. Shadowing of the signal as well as the decodability threshold of the measuring device determine the maximum distance to these observation points. Naturally, shadowing from obstacles and multipath fading effects cause the maximum observation distance to be random. The probability of observing i at point ν can be calculated from the distribution of the shadowing variable

χ and the decodability threshold ψ that must be exceeded. For the case of normally distributed shadowing studied in this paper, it is given by

$$\mathbb{P}((i,\nu,P_{\text{RX},\nu}) \in \mathcal{M}) = p(P^i_{\text{RX},\nu} \geq \psi) = \dots \quad (3)$$
$$= \frac{1}{2}\text{erfc}\left(\frac{\psi - P^i_{\text{TX}} + L_{\text{off},i} + 10\kappa\log_{10}(||\eta_i,\nu||)}{\sqrt{2\sigma^2}}\right),$$

where $\text{erfc}(\cdot)$ is the inverse error function. The sampling process can be modelled as a non-homogeneous Poisson Point Process [13] with intensity

$$\lambda(\nu) = p(P^i_{\text{RX},\nu} \geq \eta)p(\nu), \quad (4)$$

where $p(\nu)$ is the probability of conducting a measurement at point ν, and may be a degenerate random variable for planned measurement campaigns. As can be observed from (4), the intensity function is angular-symmetric, iff $p(\nu)$ is angular-symmetric and $\mathcal{A}$ is large enough. In this case, geometric localization models derive the maximum-likelihood position of the transmitter from the set N^i as their centroid

$$\hat{\eta_i} = \underbrace{\left[\tfrac{1}{v}\dots\tfrac{1}{v}\right]}_{\text{W}}\underbrace{\left[\nu_1\dots\nu_r\right]^{\mathsf{T}}}_{\text{N}^{\text{i}}}. \quad (5)$$

We omit here the proof of convergence for the sake of brevity, and refer instead to relevant literature [14]. Some authors (e.g. in [8]) propose to derive the values in W, commonly denoted as *weight vector*, through a (strictly) monotonic function $w(P^i_{\text{RX},\nu_i})$ with $\sum_{\nu\in\text{N}^i} w(P_{\text{RX},\nu}) = 1$. A common assumption is that the weight of observation points with high received signal power should be increased. The argument is that, for a small number of observations, the convergence rate of the estimator towards the true location is thereby improved. Unfortunately, the fitness of any particular weighting function has to the best of our knowledge until now not been proven within any analytical framework.

Centroid algorithms often exhibit poor performance in real-world measurement campaigns for two reasons. Firstly, the selection of measurement locations is often restricted to accessible roads or places. Hence, the necessary angular symmetry is sacrificed. This effect, known as *road bias*, is often observed in irregular measurement campaigns, e.g. in cities with a strictly regular street layout. Secondly, correlated shadowing is a further source of error in the estimation. In centroid algorithms, a skew occurs if one angular direction from the transmitter is blocked, e.g. due to a building blocking the main signal path. Weighting of measurements with the received signal power may relieve this error, yet without a proper shadowing concept the induced error due to an ill-conditioned weighting function may compromise the improved performance. This motivates our proposed hybrid localization technique which combines geometric and propagation concepts into a radio environment-aware algorithm.

IV. Proposed Propagation-Centric Localization Algorithm

In this section we propose a novel propagation-centric localization technique which explicitly integrates propagation modelling in order to improve location prediction accuracy. The key element of our approach is to explicitly consider the effect of shadowing instead of implicitly absorbing it in the general localization error, as in purely geometric-based techniques. Specifically, our algorithm makes use of additional information in the form of individual signal strength estimates, i.e. the third elements of the tuples in $\mathcal{M}$. We model propagation as a log-distance dependent attenuation combined with an AP-specific offset attenuation. The localization problem therefore becomes a joint optimization task in the spatial and propagation domains.

A. Algorithm Description

For an AP with an unknown location we first find a set of candidate positions and derive the optimum propagation parameter estimates (in the least-square sense) for each candidate position. Namely, for each location η_{cand} we need to find

$$\left(\hat{\kappa},\hat{L}_{\text{off},i}\right) = \underset{\kappa,L_{\text{off}}}{\arg\min}\sum_{\nu\in\text{N}^i}(P_{\text{TX}} - L_{\text{off}}\dots \quad (6)$$
$$-10\kappa\log_{10}||\eta_{\text{cand}},\nu|| - P^i_{\text{RX},\eta})^2.$$

We then determine the residual error, and select the point where the fitting performed best. This optimization problem is solved by running two nested optimizers as shown in the pseudocode in Fig. 1. The outer-loop optimizer firstly selects a (number of) candidate position(s) for the AP. This may be the location previously predicted by e.g. the centroid algorithm. In the fitness function of the optimizer, we derive the best fit of propagation parameters in the least-square sense. This optimization requires simple polynomial fitting with one degree of freedom. We note that the real transmit power of the transmitter does not need to be known, because an error would be absorbed into the L_{off} estimate. This is a significant advantage of our approach, since the estimation process will perform equally well without this additional information.

The residual error of the fit establishes the fitness value. The outer optimizer then iteratively tests more candidate positions. Due to the non-convex nature of the problem, evolutionary optimization algorithms are best-suited to find the error minimizing position. In our implementation, we use a Genetic Algorithm [15] for the outer-loop optimizer and a linear least square solving algorithm [16] for the inner optimizer.

B. Baseline Performance Comparison with Geometric Techniques

As a baseline comparison, we run our localization algorithm in a simulated radio environment and compare it against the centroid algorithm. Furthermore, in order to distinguish between errors arising from the propagation parameter estimation and those arising from the localization, we evaluate two versions of our algorithm: (i) our full algorithm as given in Section IV-A where radio parameters and location are jointly estimated, and (ii) a genie-aided version where radio parameters are known *a priori*. We employ the propagation model from Section II and distribute measurement points in a large regular square grid around the true location of the

procedure JOINTOPTIMIZATION($\mathcal{M}, P_{TX}, thresh$)
  $e_{\text{best}} \leftarrow \infty, \eta_{\text{best}} \leftarrow [\,], E \leftarrow [\,], R \leftarrow [\,]$
  $N \leftarrow [\pi_{(2)}(\mathcal{M})], P_{\text{RX}} \leftarrow [\pi_{(3)}(\mathcal{M})]$
  repeat ▷ Outer optimizer
    $R \leftarrow$ SELECTCANDIDATES(R,E) ▷ GA function
    for all $\eta^i_{\text{cand}} \in R$ **do** ▷ Inner optimizer
      $d \leftarrow [\,], E^i \leftarrow 0$
      for all $\nu^j \in \text{N}$ **do**
        $d^j \leftarrow 10 \log_{10} ||\nu^j, \eta^i_{\text{cand}}||$
      end for
      $(\hat{L}, \hat{\kappa}) =$LEASTSQUAREFIT($d$,$P_{\text{RX}}$)
      for all $\nu^j \in \text{N}$ **do** ▷ Error estimate
        $E^i = E^i + (P^j_{\text{RX}} - P_{TX} - \hat{L} - \hat{\kappa} d^j)^2$
      end for
      if $E^i < e_{\text{best}}$ **then**
        $e_{\text{best}} \leftarrow E^i, \eta_{\text{best}} \leftarrow \eta^i_{\text{cand}}$
      end if
    end for
  until $r_{\text{best}} < thresh$ or iterative improvement stalls
  return η_{best} ▷ Best fit location
end procedure

Fig. 1. Pseudocode of joint propagation and location estimation.$\pi_{(\cdot)}$ is the projection operator on a tuple. A^i selects the i^{th} item in the array A.

transmitter with grid resolution Δd. The shadowing standard deviation is set to $\sigma = 5\,\text{dB}$, a common value for urban deployment scenarios [12]. We set the baseline attenuation to $L_{\text{off}} = 81\,\text{dB}$, and set $\kappa = 2$, which corresponds to the free-space pathloss model for a transmitter operating at 2.4 GHz. Only observation points for which the received signal power exceeds $\psi = -95\,\text{dBm}$ are considered; the AP transmit power is set to $P_{\text{TX}} = 20\,\text{dBm}$, resulting in an average decodability distance of approximately 50 m. These values correspond to a generic IEEE 802.11 transmitter and Wi-Fi adapter, run under European ISM band regulations.

In Fig. 2 the cumulative distribution function (CDF) of the distance between estimated and true location is shown. We have run the simulation 1000 times with different realizations of the shadowing variable to acquire a sufficient sampling set. For the highest density sampling scenario with $\Delta d = 10\,\text{m}$, our propagation-aware algorithm with known radio parameters (dotted line) shows significantly better performance than the centroid-based approach (solid line). However, if radio parameters and location need to be jointly estimated (dashed line), the performance drops below that of the centroid algorithm. Note at this point that we assume uncorrelated shadowing in all directions, the best-case for the centroid algorithm. If the angular symmetry properties of the scenario are violated, as would be typically the case in a real-world measurement scenario, the centroid algorithm will necessarily perform worse. The superiority of our propagation-aware model becomes evident in scenarios with a low number of samples. The red curves in Fig. 2 show a scenario we deem more realistic for actual real-world experiments. Here, the distance between measurement points is quadrupled, resulting in an average

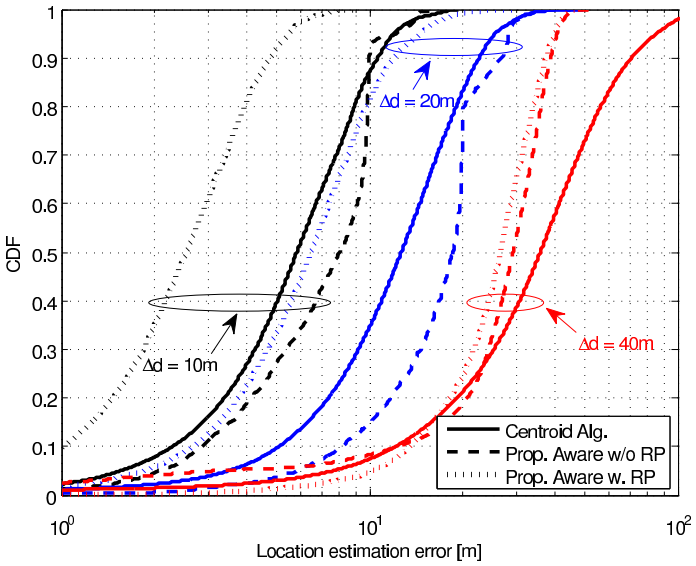

Fig. 2. Cumulative distribution function of the distance between true and estimated location of an AP. Line type determines the reported localization algorithm, line color the grid size of the measurement grid.

of 6 samples per estimation run (152 for $\Delta d = 10\,\text{m}$). Both versions of our propagation-aware algorithm outperform the centroid algorithm with an estimation error of less than 40 m compared to 70 m for 90% of the tested cases. In a scenario with $\Delta d = 20\,\text{m}$ that scales between highest and lowest density, the relative performance of the joint estimator compared to the centroid algorithm is worse than in the highest density scenario, thus a general turning point at which centroid algorithms are more feasible cannot be easily derived. The genie-aided version of our algorithm with known radio parameters is superior in all scenarios.

The results from Fig. 2 demonstrate that radio-propagation based localization exhibits significant performance gains if the propagation parameters are precisely estimated, but that joint estimation of propagation parameters and location will outperform centroid algorithms only in low-density scenarios[1]. In the following we thus focus on refining our algorithm to minimize the estimation error for the radio parameters when the location of a subset of APs is known. The acquired propagation information is then fed as an input to the localization process for the remaining APs whose location is unknown.

V. RADIO ENVIRONMENT PARAMETER ESTIMATION WITH A SINGLE KNOWN AP

In this section we derive the error bounds of the estimation of the radio propagation parameters from measurements with a single AP. For the sake of clarity, we will for the moment omit the transmitter index in our equations. As discussed in Section IV-A, the inner optimizer applies a linear regression to retrieve the (L_{off}, κ)-tuple for the respective AP. The

[1]We emphasize however that low-density spatial sampling is representative of data gathered from practically feasible large-scale measurement campaigns for which these localization methods are intended.

confidence intervals of this process can be semi-analytically determined if the true location of the AP is known. The α quantile of the parameters is determined as

$$\hat{L}_{\text{off}} = L_{\text{off}} \pm t(\alpha) \times s_L \quad (7)$$

$$\hat{\kappa} = \kappa \pm t(\alpha) \times s_\kappa \quad (8)$$

where $t(\cdot)$ is the inverse of the CDF of the student-t distribution with $|\mathrm{N}| - 2$ degrees of freedom, and s_L and s_κ are parameters as given below. We define

$$d = |\mathrm{N}|^{-1} \sum_{\nu \in \mathrm{N}} \log_{10} ||\eta, \nu|| \quad (9)$$

as the *mean log-distance of the observation points*, and

$$m = |\mathrm{N}|^{-1/2} \times \ldots \left(\sum_{\nu \in \mathrm{N}} \left(P_{\mathrm{TX}} - \hat{L}_{\text{off}} - 10\hat{\kappa} \log_{10} ||\eta, \nu|| - P_{\mathrm{RX},\nu} \right)^2 \right)^{1/2} \quad (10)$$

as the *root mean-square error (RMSE) of the inner-loop optimizer*. Applying simple findings from regression analysis theory [17], we find that

$$s_L = m \times \sqrt{\frac{1}{|\mathrm{N}|} \left(1 + \frac{|\mathrm{N}| \times d}{10 \sum_{\nu \in \mathrm{N}} (\log_{10} ||\eta, \nu|| - d)^2} \right)} \quad (11)$$

and

$$s_\kappa = m \times \sqrt{\frac{1}{\sum_{\nu \in \mathrm{N}} (10 \log_{10} ||\eta, \nu|| - d)^2}}. \quad (12)$$

It is straightforward to show that the second factor in (11) and (12) is constant and solely determined by the geometry of the measurement scenario. With this knowledge, we can optimize the selection of measurement point locations by *maximizing the log-distance variance*. Note at this point that this geometric optimization is strictly limited within the boundary of the decodability range.

The RMSE m reflects the shadowing component inherent in the data. For a shadowing-free environment, it takes a zero value and the confidence interval collapses towards the true parameters of the shadowing distribution. The RMSE is, like the shadowing, a random variable with a sum-lognormal distribution. Unfortunately, this distribution is analytically intractable, but several approximations exist [18]. Numerically, we find that for non-zero shadowing, the average RMSE is upper-bounded by the shadowing standard deviation. Furthermore, as shown in Fig. 3, the RMSE approaches the value of the shadowing deviation for a large number of samples.

The poor performance of our joint estimation algorithm for high sample-size scenarios that we demonstrate in Fig. 2 can thus be explained by the divergence of the confidence interval. Whereas the RMSE converges towards the shadowing standard deviation, the geometry factor in (11) and (12) is bound by the decodability threshold of the system and is constant. The inverse CDF of the student-t distribution monotonically decreases and finally converges towards a fixed

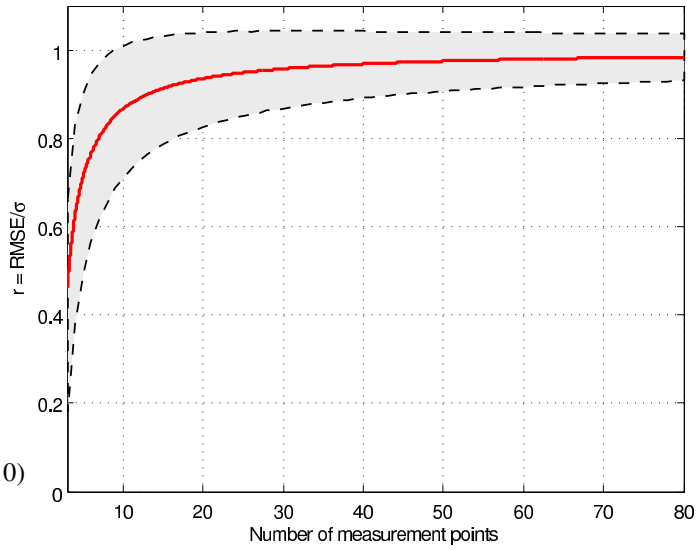

Fig. 3. Ratio of the root mean-square error and the shadowing deviation σ for single AP scenario with known location vs. number of measurement points. The red center line shows the average ratio. The gray area shows the 25^{th} and 75^{th} percentile after 100,000 simulation runs.

(non-zero) value for $|\mathrm{N}| \to \infty$. Hence, whereas for centroid algorithms the estimator converges towards the true center of all measurement points (the ideal case for uncorrelated shadowing), propagation-aware algorithms will always exhibit a non-declining error.

Importantly, this finding further supports the argument that the correct radio parameter estimation is a key challenge for the determination of the location of a transmitter. A joint estimation of the radio environment and the location of an AP shows at best marginally better performance than the centroid estimation, but with the knowledge of the propagation parameters the localization precision can be significantly improved. In the following section we therefore propose a mixed approach, where we combine measurements from multiple APs in order to improve propagation prediction accuracy.

VI. Radio Environment Parameter Estimation with Multiple Known APs

Our initial analysis in Section IV-B revealed that, for the ideal case of angular-symmetric shadowing, simpler geometric approached outperform our algorithm when radio parameters and location need to be jointly estimated. However, we have also shown that providing *a priori* knowledge of radio parameters to our propagation-centric localization algorithm always yields superior performance compared to conventional geometric algorithms. In Section V we derived the error bounds of such parameter estimation for a single AP. In this section we will investigate how to combine parameter estimations from *multiple* APs in order to further increase radio parameter estimation accuracy for subsequent localization of the remaining unknown APs. We propose a method to combine sample sets to derive a global pathloss coefficient κ estimate, since only this parameter is assumed to be the same for all

transmitters (if area $\mathcal{A}$ is sufficiently small). The advantage of this approach is that subsequent localization tasks will only need to estimate one radio parameter for each unknown AP i, namely the local offset $\hat{L}^i_{\text{off}}$.

A. Proposed Algorithm for Estimating the Global κ Parameter from multiple APs

Our proposed algorithm works as follows. In order to find the global κ value for the set of known APs, we first derive individual $(\hat{L}^i_{\text{off,single}}, \hat{\kappa}^i_{\text{single}})$ estimates. We subsequently build a new set of *normalized* pathloss estimate tuples $\mathcal{R} = \{(L^i_{\text{norm},\nu}, ||\nu, \eta_i||)\}$ by calculating for all known APs and their observation locations $\nu \in \mathrm{N}^i$

$$L^i_{\text{norm},\nu} = P^i_{\text{TX}} - P_{\text{RX},\nu} - \hat{L}^i_{off}. \tag{13}$$

In the final step of our algorithm we apply polynomial fitting to find the new global κ estimate,

$$\hat{\kappa}_{\text{est,combined}} = \arg\min_{\kappa} \sum_{t\in\mathcal{R}} \left(\pi_{(1)}(t) - 10\kappa \log_{10} \pi_{(2)}(t)\right)^2, \tag{14}$$

where $\pi_{(\cdot)}$ is the projection operator. This algorithm removes the local offset estimated in the first step and builds a new $\hat{\kappa}_{\text{est,combined}}$ estimation solely based on the combined measurements from the known APs.

B. Performance Comparison with Single AP and Averaging Methods

In Fig. 4 we show the standard deviation σ of the global κ estimation for the proposed algorithm if we include a different number of APs into our estimation process. We postulate σ to be equivalent to the RMSE due to the properties of the developed unbiased estimator. Fig. 4 thus shows the extent of error in the global estimator as derived from 100,000 simulation runs per the selected number of APs. For each AP we have chosen five distances from the AP in the range $[10\,\text{m}, 30\,\text{m}]$ and normalized the shadowing term so that the deviation in local $\hat{\kappa}^i_{\text{single}}$ estimates corresponds to the given deviation value. Shadowing standard deviation and true κ are selected as in Section IV-B and individual L_{off} values are i.i.d. random variables with $L^i_{\text{off}} \sim \mathcal{N}\left(81\,\text{dB}, (10\,\text{dB})^2\right)$. For a baseline comparison we show in blue the RMSE for a $\hat{\kappa}$ with a single AP as proposed in Sectionr̃adio-environment-estimation-single-ap, *ceteris paribus*.

We note that the error in the multiple AP estimator exceeds the single AP case. Multiple AP estimator results are comparable to the single AP equivalent only for a larger number of transmitters (e.g. if the local κ estimation RMSE is 0.1, it requires five measurements each from 19 APs to achieve an equivalent performance as with a single AP with 7×5 measurements). The reason for this additional error lies in the normalization step of the algorithm. Due to the estimation error in the local offset L_{off}, the algorithm introduces an unwanted offset into the calculations. This results in a systematic error for subsequent estimation steps; better knowledge of the local offset can alleviate this problem.

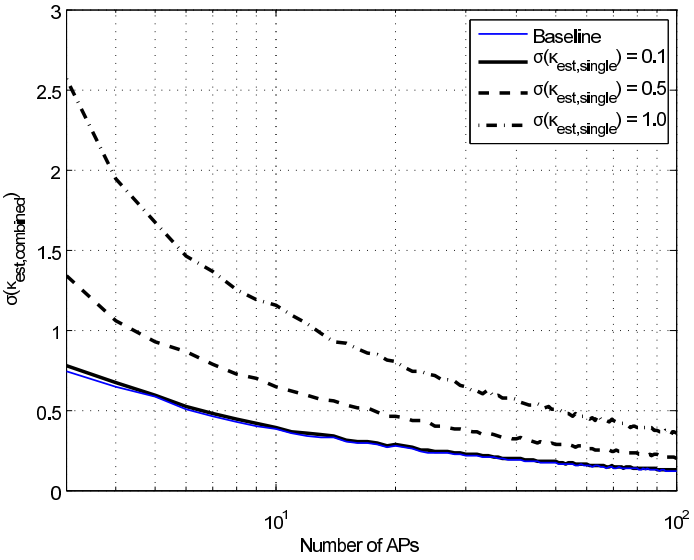

Fig. 4. Root mean square error of the global pathloss coefficient κ estimation if estimation errors are present in its local estimation. The graph shows the global RMSE for different levels of local RMSE. The local κ is estimated using 5 samples per AP at distances uniformly distributed between 10 m and 30 m ($\kappa = 2, \sigma\left(L_{\text{off},i}\right) = 10dB$) with the stated error. The baseline for comparison is a single AP estimation of the pathloss coefficient, with the equivalent number of samples as the total in the multiple AP case.

The question arises whether the proposed algorithm is thus superior to a simple averaging over κ estimates of multiple APs. The pathloss estimator is unbiased and a simple averaging estimator eventually converges towards the true κ value of the scenario. In order to investigate this assumption, we have repeated our simulations, this time comparing local κ averaging with our global normalization. In Fig. 5 we show the standard deviation of the global estimation if local κ estimates are averaged over the standard deviation of our normalization method. Values below the diagonal blue line indicate superior performance of the normalization algorithm. We find that for a reasonable number of APs, the error in the estimates converges faster with normalization than with averaging. With 50 APs, the standard deviation of the estimate is decreased from 1 to 0.5 using normalization instead of averaging , and for a scenario where the standard deviation for the averaging method is 1.5 using normalization with 100 APs it is lowered to 0.5. Our proposed normalization method to combine samples from multiple APs is therefore superior for most real-world measurement scenarios.

VII. Performance of Localization Algorithms on Data from a Measurement Campaign

In this section we evaluate our proposed localization algorithm using data from a measurement campaign of Wi-Fi APs, in order to validate it effectiveness for deriving the spatial structure of opportunistically deployed real-world networks.

A. Measurement Campaign Description

We have carried out a small-scale Wi-Fi penetration study in the small town of Mehrhoog, Germany, which is located on

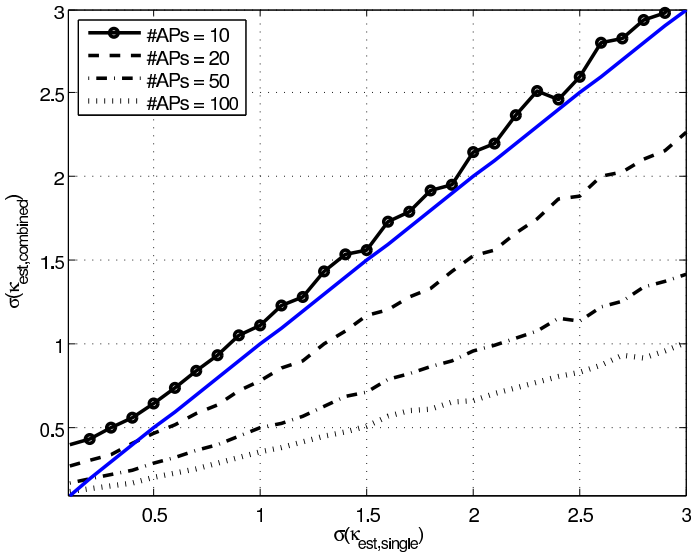

Fig. 5. RMSE of κ estimation: Comparison between RMSE for κ estimation through averaging over the locally estimated values and κ estimation through normalization. With small number of routers the κ estimation error lowers below the compared singular variance, marked by the blue isoperformance line.

flat terrain and is surrounded by a large agricultural area. These properties make this 6000-citizen residential area an optimal test case for an isolated and relatively homogeneous radio environment. The campaign was conducted during two days in July 2012 under static weather conditions. Our test setup consisted of four off-the-shelf Alfa Network AWUS051NH 802.11 Wi-Fi adapters we connected to three directional panel antennas with 120 degree horizontal beam and 11 dBi gain, and one omni-directional antenna with 5 dBi gain. For self-localization we used a Garmin GPS receiver with DGPS support (vertical precision $\leq 2\,\mathrm{m}$). A wooden mast to which the antennas were mounted was manually titled to maintain the polarization plane and the omnidirectional antenna was located 3 m above ground. Measurements were conducted at 217 measurement points (MPs) with an approximate inter-point distance of 5 m. To eliminate fast fading, we averaged over 60 signal estimates in a one-minute measurement period. We focus here on the data retrieved from the omnidirectional antenna and defer reporting on the results from the directional measurements to a separate publication.

B. Localization Algorithm Performance Analysis

During the campaign we discovered 368 Wi-Fi APs in an area of $580\,\mathrm{m} \times 740\,\mathrm{m}$, i.e. an average density of 857 APs/km^2. Compared to earlier studies [4] we found that the residential AP density has further increased, in line with the wider availability of broadband internet access for home users. By mapping Wi-Fi identifiers to building owners, we were able to reliably locate 9 user-deployed APs in the studied area. These APs constitute the set of known transmitters for the analysis.

In Table I we summarize the characteristics of the APs and results we retrieved when running our proposed algorithms with the measurement campaign data[2]. Using single AP parameter estimation, for 4 out of the 9 APs we estimate unrealistically low pathloss coefficients with values below 1.1. The results are indicative of a high shadowing-induced estimation error.

Table I also shows the performance of two variants of the centroid algorithm, which we use as our benchmark. In the unmodified centroid algorithm, all measurement points are weighted equally. We select the weighting function $w(\cdot)$ as

$$w(P^i_{\mathrm{RX},\nu_j}) = (20 - P^i_{\mathrm{RX},\nu_j})^{-1} \sum_{\nu \in \mathrm{N}^i} \frac{1}{20 - P^i_{\mathrm{RX},\nu}} \qquad (15)$$

to increase the weight of high-RSSI measurement points based on the pathloss (assuming $P_{\mathrm{TX}} = 20\,\mathrm{dBm}$). Both algorithms exhibit a high variance in their localization performance. AP #2 with the highest variance in measurement locations is predicted to be located more than 90 m from its real location, while AP #1 is precisely identified. Analyzing the individual geometries of these APs, we found that a high road bias is apparent for AP #2. Here, the strongest signal path is guided by the main road on which the measurements were conducted.

Our localization algorithm with known local radio parameters on average performs better than geometric localization in the studied cases. In particular for APs for which the centroid algorithms suffer from a high bias, our propagation-centric localization gives better estimates. On the other hand, if centroid algorithms benefit from a well-aligned, i.e. near circular, measurement geometry, they are superior to our proposed technique even with the additional genie-aided information. If localization and radio parameter estimation are jointly conducted, centroid algorithms and our approach perform similarly. We note though that the average error is lower for our propagation-centric approach, in particular for APs where the centroid-based estimates are highly off. Given that the aim of the campaign is to improve the average localization error for large sets of APs, we therefore argue that our propagation-centric algorithm is overall superior.

Finally, we derive global κ estimates through a subset of the known APs to estimate the location of one of the APs not in the training set, based on the algorithm proposed in Section VI. The last row set in Table I lists the localization errors using this algorithm, where values in brackets indicate the standard deviation for all possible AP combinations of the sample set. We observe that this method largely improves the localization performance and estimates the AP locations with high accuracy. In the degenerate case of AP #2, where neither centroid algorithms nor local parameter estimations perform well, a precision of 22 m can be achieved using the global κ estimation even with a small training set. This indicates that

[2]We note that the anomalous data for AP #7 (an almost 20 dB lower local offset) is likely due to this AP running a modified firmware that would allow a 12 dB higher transmit power than allowed by European regulators. European regulations allow operating IEEE 802.11 access points at a maximum EIRP of 20 dBm. Updates of the firmware, e.g. with open-source software, often remove this limitation and increase the power level to 32 dBm as permitted in the US.

TABLE I
LOCATION ESTIMATION ERROR FOR MEHRHOOG TEST SCENARIO

	Access Point ID								
Measurement characteristics	#1	#2	#3	#4	#5	#6	#7	#8	#9
no. of measurements	9	10	8	4	10	21	11	7	12
σ log-distance MP$\rightarrow$AP	6.3	17.7	10.4	9.3	7.7	9.8	6.3	6.7	5.6
estimated local κ	0.55	1.81	1.08	0.99	1.95	1.02	3.43	2.28	2.30
estimated local L_{off}	92.5	65.8	84.9	90.5	65.3	84.3	47.2	68.7	65.1
Reference localization errors									
centroid alg.[m]	6.6	96.1	33.5	13.6	41.6	57.2	15.5	27.0	24.7
weighted centroid alg.[m]	6.5	91.2	32.5	13.4	39.6	56.9	13.7	25.7	23.9
local known κ alg. [m]	16.3	23.6	22.3	14.7	23.3	13.5	11.7	5.1	6.1
Local κ estimation									
localization error* [m]	16.4 (0.0)	171 (103)	42.2 (0.6)	33.7 (0.1)	25.6 (1.4)	48.2 (0.3)	12.4 (0.1)	14.1 (0.4)	10.9 (0.2)
Global κ estimation									
localization error†, 5 APs [m]	8.9 (5.2)	22.1 (2.2)	22.5 (10.6)	15.5 (3.8)	22.6 (0.7)	40.3 (10.2)	10.0 (0.1)	12.7 (1.9)	15.0 (2.7)
localization error†, 6 APs [m]	8.8 (5.2)	21.6 (1.0)	21.9 (9.3)	14.3 (3.5)	22.4 (0.2)	43.3 (3.0)	10.0 (0.1)	12.7 (1.3)	15.0 (2.3)
localization error†, 7 APs [m]	7.3 (5.0)	21.5 (0.1)	21.1 (9.0)	13.2 (2.8)	22.5 (0.3)	43.7 (1.6)	10.0 (0.1)	13.6 (0.8)	16.0 (1.5)
localization error†, 8 APs [m]	3.1 (-)	21.5 (-)	8.9 (-)	11.5 (-)	22.2 (-)	43.9 (-)	10.0 (-)	13.7 (-)	16.2 (-)

* Values in parentheses show RMSE after 500 runs of optimizer.
† Values in parentheses show RMSE using all possible combinations.

using a limited set of known APs, our proposed propagation-centric localization method with global radio parameter estimation would consistently provide better performance than centroid localization algorithms for estimating the location of unknown APs from a large-scale measurement campaign.

VIII. CONCLUSIONS

In this paper we have proposed a novel propagation-centric localization method to derive the spatial structure of opportunistic wireless networks through measurement campaigns with simple COTS equipment. The proposed method combines the geometry of observation points and a jointly estimated radio propagation model, in order to improve the overall accuracy of node location estimation compared to purely geometry-centric approaches. Our algorithm is, to the best of our knowledge, the first proposed technique which integrates measurements from a limited subset of known APs through joint propagation modelling to improve the localization of a large set of unknown APs. Through theoretical studies, simulations, and a measurement campaign we have shown that our proposed radio environment-aware algorithm achieves a significantly reduced localization error compared to conventional localization approaches, in particular for cases where angular asymmetries in the observation point distribution exist, as is the case for large real-world measurement campaigns. We emphasize that our method is widely applicable for real measurement campaigns due to non-stringent requirements of COTS equipment and information about only a few known AP locations.

REFERENCES

[1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.

[2] J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 156–163, Nov. 2010.

[3] M. Haenggi, *Stochastic Geometry for Wireless Networks*. CUP, 2012.

[4] K. Jones and L. Liu, "What Where Wi: An Analysis of Millions of Wi-Fi Access Points," in *IEEE International Conference on Portable Information Devices (PORTABLE07)*, 2007, pp. 1–4.

[5] L. Driskell and F. Wang, "Mapping digital divide in neighborhoods: Wi-Fi access in Baton Rouge, Louisiana," *Annals of GIS*, vol. 15, no. 1, pp. 35–46, 2009.

[6] C. Phillips, R. Senior, D. Sicker, and D. Grunwald, "Robust Coverage and Performance Testing for Large-Area Wireless Networks," in *AccessNets*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, C. Wang, Ed. Springer, 2009, vol. 6, pp. 457–469.

[7] P. Mähönen, J. Riihijärvi, and A. Kivrak, "Statistical characterization of transmitter locations based on signal strength measurements," in *5th IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, May 2010, pp. 283–288.

[8] A. Tsui, W.-C. Lin, W.-J. Chen, P. Huang, and H.-H. Chu, "Accuracy Performance Analysis between War Driving and War Walking in Metropolitan Wi-Fi Localization," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1551–1562, Nov. 2010.

[9] A. Subramanian, P. Deshpande, J. Gaojgao, and S. Das, "Drive-by localization of roadside wifi networks," in *The 27th IEEE Conference on Computer Communications (INFOCOM)*, April 2008, pp. 718–725.

[10] Y. Chen, Z. Liu, Y. Ding, and X. Fu, "On sampling signal strength for localization using a directional antenna," in *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'11)*, June 2011, pp. 1–6.

[11] J. Koo and H. Cha, "Unsupervised Locating of WiFi Access Points Using Smartphones," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. PP, no. 99, pp. 1–13, 2012.

[12] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Wiley, 2002.

[13] R. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*, 1st ed. Springer, 2010.

[14] V. Voinov and M. Nikulin, *Unbiased Estimators and their Applications: Volume 2: Multivariate Case*, 1st ed. Springer, 1996.

[15] K. Deb, *Multi-objective Optimization using Evolutionary Algorithms*. Wiley, 2004.

[16] A. Björck, *Numerical Methods for Least Squares Problems*. Siam, 1996.

[17] G. Vining, E. Peck, and D. Montgomery, *Introduction to Linear Regression Analysis*, 5th ed. Wiley, 2012.

[18] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Estimating the distribution of a sum of independent lognormal random variables," *IEEE Transactions on Communications*, vol. 43, no. 12, p. 2869, 1995.