International Conference on Indoor Positioning and Indoor Navigation (IPIN), 24-27 September 2018, Nantes, France

A Bayesian Approach to Dealing with Device Heterogeneity in an Indoor Positioning System

Kyle F. Davies*†, Ian G. Jones† and Jonathan L. Shapiro*
*School of Computer Science
University of Manchester
Manchester, UK
Email: {kyle.davies, jonathan.l.shapiro}@manchester.ac.uk
†Purple WiFi Ltd
Oldham, UK
Email: ian.jones@purplewifi.com

***Abstract*—There are many practical applications in which the ability to localize devices such as phones, tablets, and mobile equipment is important. One of the issues which makes this difficult is the fact that devices are all different, so an approach which is robust against device heterogeneity would be an advance. In this paper, a method for estimating the positions of transmitting devices using Wi-Fi and a network of access points (APs) is proposed and investigated. The APs can also function as transmitters; as such the method allows simultaneous calibration and localization, so no fingerprinting or separate calibration is required. A hierarchical Bayesian probabilistic model is used with separate but conditionally-related parameters for each transmitter and receiver to tackle the device inhomogeneity problem. The output is a probability distribution over the location of each device from which the expected location and measures of uncertainty in location can be obtained. The system was implemented in an office environment using heterogeneous transmitters and receivers. The system localized the devices with a median error of 1.7 meters and within 4.32 meters with 95% confidence. We discovered that it is more important to account for inhomogeneity in the transmitters than in the receivers. Removing the former from the model results in a median error of $6.57\,\mathrm{m}(10.56\,\mathrm{m})$ whereas removing the latter results in a median error $1.93\,\mathrm{m}(4.64\,\mathrm{m})$. We argue that the technique could be used to cope with other types of inhomogeneities in the environments or the Wi-Fi equipment.**

I. Introduction

The widespread adoption of GPS-capable mobile devices has enabled the development of mature, valuable outdoor location-based services. Similarly, indoor positioning systems can enable a range of applications such as asset-tracking, navigation, and context-aware marketing, the potential of which has driven research. However, the challenging nature of complex indoor environments has prevented any technology from fully solving the indoor positioning problem.

A wide variety of location-sensitive signals have been used to solve the problem, such as ultra-wideband, visual, ultrasonic, and magnetic [1]–[4]. Each of these has their own advantages and drawbacks. A popular approach, motivated by the possibility to leverage previously deployed infrastructure and the ubiquity of the technology in mobile devices, is the use of measurements of Wi-Fi signals made by access points (APs).

There are several different aspects of Wi-Fi signals that can be exploited to pinpoint their origin, such as received signal strength (RSS), channel-state information, time-of-flight, and antenna phase difference. Additionally this information can be used in either a fingerprinting framework or with a model-based approach. Due to its ubiquitous availability in commodity hardware, in this paper we present an RSS-based positioning system designed to address two main issues that have presented difficulties to previous methods: the need for onerous system calibration (fingerprinting or war-driving), and diversity in the transmission and reception characteristics of devices (the issue of device heterogeneity).

II. Related Work

A. Self-calibrating positioning systems

The most accurate Wi-Fi based indoor positioning systems have typically required extensive site survey to measure the radio environment: the fingerprinting

978-1-5386-5635-8/18/$31.00 ©2018 IEEE

approach [5], [6]. However, the ongoing effort required to produce and maintain the fingerprint database is a significant barrier to widespread adoption of these methods. Therefore there has been interest in developing calibration-free methods, which do not require such effort [7]. Of particular interest are self-calibrating approaches, which do not require manual calibration effort, but still adapt to specific radio-propagation environments.

Notable examples of self-calibrating systems include Triangular Interpolation and eXtrapolation (TIX) [8], the Signal-Distance Map (SDM) technique [9], and the dynamic mapping system presented in [10]. Each of these has in common the idea of using inter-AP RSS measurements to automatically adjust the positioning algorithm to fit the environment.

The methods listed above all work with point estimates of the radio propagation model. A different approach, similar to ours, is to build a Bayesian hierarchical model based on a path-loss model for RSS, such as that presented in [12]. In such a set up, all unknowns are treated the same, and a posterior distribution over them accounting for measured data is calculated. For a positioning system, the unknowns of interest are the device locations, but the model can be expanded to include extra explanatory variables; for example, the model presented in [12] includes per-AP slopes and intercepts for the path-loss equation.

None of the listed methods deal with diversity in device transmission characteristics. In a self-calibrating system, this may be expected to be of particular importance, as APs can make use of permanent power sources and high-gain antennas to produce a higher transmit power. Calibration based on measurements between APs may be inappropriate when applied on mobile devices due to this effect. Therefore a focus of our work was in mitigating the effects of this diversity.

B. Dealing with device diversity

Positioning systems based on RSS measurements work by exploiting the relationship between signal strength and physical position. However, this relationship is disrupted because Wi-Fi devices do not all transmit signals at the same power level. This can be caused by, for example, amplification level, antenna design, and other hardware differences. This issue is of particular importance in self-calibrating systems, as the transmission characteristics of the APs are likely to be significantly different from those of mobile devices. There have been two main approaches to compensating for the problem of device diversity: device mapping, and robust feature design [13].

The device mapping approach attempts to correct measured RSS values using a database of Wi-Fi device characteristics [14], [15]. This approach is practically infeasible in general, due to the huge range of hardware available on the market, as well as the difficulty of determining what model an unknown device is. It has also been reported that significant differences in transmission power can be seen even with identical device models [15].

The other main approach to dealing with device diversity has been replacing raw RSS readings with features designed to be robust to differences in transmit power. For example, the RSS values from individual sensors can be replaced with differences of signal strength between different sensors, resulting in features which are invariant to the power of the transmission [16].

A learning-based approach was used in the WiGEM system [17], which modelled RSS values as a Gaussian mixture. Mixture components were indexed by two latent variables, which were interpreted as representing position and transmission power level. However, this approach has several drawbacks. Firstly, the identification of the latent variables with position and power is only enforced by the mixture component initialization, and need not hold as model fitting progresses. Secondly, the position and transmission power variables are discretized, meaning that a finer granularity or larger space requires more mixture components. Finally, the learned mixture is only valid for a single device and learning procedure must be repeated for each tracked device.

We will now describe our approach to the indoor wireless positioning problem, in which the issue of device diversity is dealt with by including unobserved variables representing the transmission power of each device, which are learned as RSS data are collected.

III. METHOD

A. Overview

The situation is an environment in which there are a set of receivers or sensors in fixed and known locations. Also in the environment are many devices which are intermittently transmitting. These could be mobile devices such as phones, tablets, or movable equipment, or could be stationary. The collection of transmitters can change over time, such as when people come in and out of the environment. The receivers can also function as transmitters which are in known locations. The goal is to determine a good approximation to the unknown location of each device.

We take a generative, hierarchical Bayesian approach. A Bayesian approach means that all unknown quantities including the location are treated probabilistically with prior probabilities. That it is generative means that the starting point is a hypothesis of a probability function which *generated* the data, i.e. the likelihood probability function $P(\text{data}|\text{variables, parameters})$. That it hierarchical means that the parameters of the priors of the Bayesian model are themselves treated as random variables with their own (hyper)priors. This allows the introduction of inhomogeneity while preserving a similarity between devices. What is novel about this work is that each device has a independent parameter governing its transmission power or gain. Thus, the model copes with device heterogeneity, but at the expense of greater model complexity.

The output of our model will be a probability distribution over possible locations. The mean of this distribution can be used as an estimate of the location, but also measures of uncertainty in the location can be found from covariances or confidence or credible regions. The alternative approach, called point prediction, is to find the best estimate of location without measures of uncertainty. Point prediction can be computationally more simple, depending on how it is done, but knowledge of how well the location is known is important in many situations. For example, it could be used to provide probabilities of a device being in one room versus another.

B. The Model

The model is based on the well-known Log-Distance Path Loss (LDPL) model from standard RF propagation theory, in which the power received at a device labelled i from device k is given, in dBm, by

$$s_{ik} = r_i + t_k - 10\gamma_i \log_{10} \frac{d_{ik}}{d_0} + \epsilon_i, \tag{1}$$

where γ_i is the path-loss exponent, which we have allowed to depend on the receiving device, d_{ik} is the distance between transmitter and receiver, and d_0 is a reference distance (treated as $1\,\mathrm{m}$). The noise term, ϵ_i, is assumed to be normally distributed with zero mean and variance σ^2 and is assumed independent for all receivers i. The power received at the reference distance d_0 is given by $t_k + r_i$, where the first variable is a property of the transmitter, and the latter is a property of the receiver. These variables allow us to model the effects of device diversity in both the receiver and transmitter.

What is known in advance includes the location of all receivers and some of the transmitters (those which are also receivers) and the value of d_0. Observed are the received signals s_{ik}^{τ} at each receiver i from each transmitter k at time bin τ. The identity of the transmitters (e.g. MAC addresses) are also observed. Let N, M, D be the number of receivers, transmitters, and readings respectively. Both transmitters with known and with unknown locations are included in M.

The goal is to infer the locations of all transmitters with unknown locations $\mathbf{x_k}$, which are treated as three-dimensional. The other quantities will be treated as so-called "nuisance variables" and should be integrated out at the end. The transmitters at known locations are used to simultaneously calibrate the model.

Given (1), the fact that d_{ik} is the Euclidean distance between the position of receiver i and transmitter k, and the distribution of noise ϵ_i, it follows that the conditional distribution $P(s_{ik}^{\tau}|r_i, t_k, \gamma_i, \mathbf{x_i}, \mathbf{x_k^{\tau}})$ is proportional to a normal distribution with mean,

$$\mu_{ik}^{\tau} = r_i + t_k - 10\gamma_i \log_{10} \frac{\|\mathbf{x_i} - \mathbf{x_k^{\tau}}\|_2}{d_0}, \tag{2}$$

and standard deviation σ. Here τ is an index over time, and indices i for receivers and k for transmitters will be consistently used. The Euclidean distance is the usual vector 2-norm, $\|\vec{y}\|_2 = \sqrt{\sum_{j=1}^{d} y_j^2}$ for a d-vector $\vec{y}$.

Of course, to infer the position of any transmitter k, the values of all observed receivers will be needed. We use the notation that vector, $s_{1k}, s_{2k}, \ldots, s_{Nk}$ is written as as a set $\{s_{ik}\}|_{i=1}^{N}$. Under the assumption that the noise is independent on each receiver and at each reading, it follows that the probability of the set of all observables, $\{s_{ik}^{\tau}\}\,|_{i=1}^{N}|_{k=1}^{M}|_{\tau=1}^{D}$, given the other variables and parameters, called the likelihood, is the product over normal distributions as described above,

$$\mathcal{L} = \prod_{\tau}^{D} \prod_{k}^{M} \prod_{i}^{N} \mathcal{N}(s_{ik}^{\tau}|\mu_{ik}^{\tau}, \sigma). \tag{3}$$

where $\mathcal{N}(y|m, \sigma)$ denotes a normal distribution as a function of y with mean m and standard deviation σ.

Using Bayes rule and the priors described in Table I, one can in write down the posterior distribution up to an unknown normalization factor, because the posterior is proportional to the product of the likelihood and the priors. The posterior is the probability of the unknown location given the receiver readings, the known locations, and the other variables and parameters.

We cast the model into a hierarchical Bayesian form, to share information about the unknown parameters

TABLE I
THE PRIORS AND HYPERPRIORS OF THE MODEL

Variable	Prior
$\boldsymbol{x}$	$\mathcal{N}(\boldsymbol{\mu}_x, \Sigma_x)$
t_k	$\mathcal{N}(\mu_t, \sigma_t^2)$
μ_t	$\mathcal{N}(0, 10)$
σ_t^2	$\mathrm{HalfCauchy}(10)$
r_i	$\mathcal{N}(0, \sigma_r^2)$
σ_r^2	$\mathrm{HalfCauchy}(10)$
γ_i	$\mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$
μ_γ	$\mathcal{N}(2, 4)$
σ_γ^2	$\mathrm{HalfCauchy}(10)$
σ^2	$\mathrm{HalfCauchy}(10)$

t_k, r_i, γ_i between devices, as from physical considerations we expect them to be similar. To explain this, take the example of one set of variables $\{t_k\}|_{k=1}^{M}$, which are associated with the transmitters. Because each transmitter could be different, they are given individual sets of prior parameters. However, they share a hyperprior probability from which prior parameters are drawn.

Let θ_k^t be the parameters of the prior associated with t_k. In this case θ_k^t represents the mean and variance of a normal distribution. Let ϕ^t be the parameters the shared hyperprior for all the t variables. Then the priors can be generated by sampling from the hyperpriors, and variable t_k and $t_{j\neq i}$ are correlated through the shared hyperparameters ϕ^t. I.e. they are only conditionally independent. The two-layer hierarchy is shown in Fig. 1

The full model graph is shown in Fig. 2. This representation is sometimes called a boilerplate representation. The enclosing rectangles, called plates, represent replication of the variables they enclose. The black plate shows replication over receivers i; the magenta plate represents replication over the transmitters k. The dashed plate shows replication over readings. Using the example of the γs, it can be seen that these are replicated N times, one for each receiver, and they feed into the mean of the receiver sensor reading. Each γ has its own prior parameters, but they share the hyperparameters, because these are outside the plate. This graph would be very complicated without the use of this graphical model notation. This notation is in wide within the data analysis community, at least since Buntine [18].

IV. INFERENCE

Required is a distribution over the location of the devices of interest given the relevant knowledge, and marginalizing over nuisance variables. This distribution is called the *posterior*. Using Bayes rule and the priors listed in Table I, the useful distribution can be found up to proportional factor by writing (3) explicitly in terms of the desired location $\vec{x_k}$, multiplying by the priors, and integrating out the nuisance variables.

However, as is typically the case in Bayesian modelling, exact calculation of this probability function is not possible because the integrals are intractable, including the normalization integral and the integrals required to remove dependence on the nuisance variables. Therefore, we use an approximate method called Markov Chain Monte Carlo (MCMC), which is very widely used in Bayesian modelling. (For a pedagogical introduction, see Chapters 11 and 12 of “Bayesian Data Analysis” [19].) The basic idea of MCMC is to produce a Markov Chain which visits points in the space of all variable values in proportion to their probability in the posterior. If run long enough, the result is a set of points, although correlated through the Markov process, approximate a set of independent samples from the posterior distribution. Each sample contains values for all variables: the unknown locations, the γs, ts, rs and the θs. Thus, if a location $\vec{x_k}$ is averaged over a set of samples, it gives an approximation to the average location of transmitter k, with the other variables integrated out. Other statistics about the location could also be estimated, such as the standard deviation or quartiles by computing empirical values over the samples.

In particular, we use Hamiltonian Markov Chain Monte Marlo (HMC) as implemented in the Python package PyMC3 [20]. Hamiltonian MCMC, introduced in the context of statistical inference by Neal [21], increases the rate of sampling through the variable space.

V. EVALUATION

In order to evaluate the proposed model, we collected RSS data using commodity Wi-Fi APs (Openmesh OM2P). Two sets of APs were deployed: one set of 16 using internal on-board antennas, and another set of 15 which were outfitted with external antennas. Readings were taken from a set of 5 Android mobile devices (Samsung S4, S6, S7; Google Pixel; NVidia Shield K1), which were placed at 15 known locations in a typical office environment (floor area $305\,\mathrm{m}^2$) and configured to periodically broadcast Wi-Fi signals. The APs were transmitting signals concurrently, the signal strengths of which were also captured. The RSS values for all APs were aggregated into a single measurement vector for each transmitter MAC address and time bin by taking

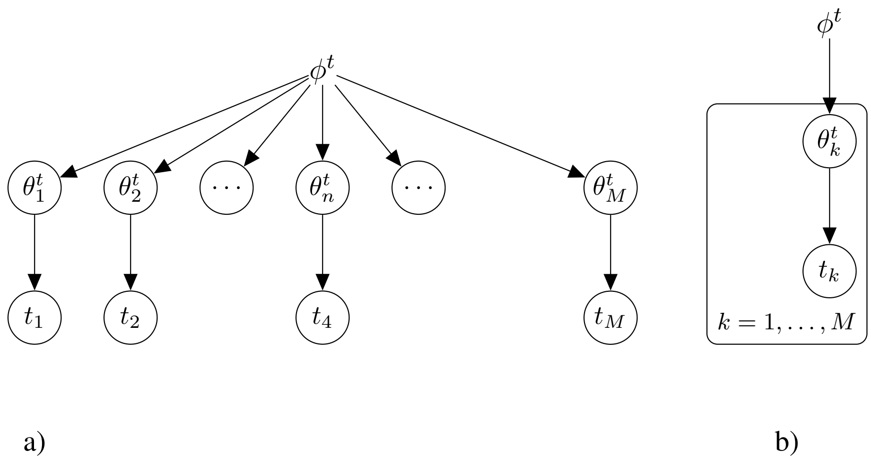

Fig. 1. A piece of the hierarchical model involving transmitter constants t_k. Each individual transmitter has its own prior parameters θ_k^t, but these are all sampled from the same hyperprior with the same hyperparameters ϕ^t. Note: θ_k^t and ϕ^t can each represent all parameters of a prior, e.g. mean and variance. The circles represent random variables; arrows express dependencies; symbols not enclose in a shape are constants. a) An example of the hierarchical nature of parameter dependence. b) The same model in a) but using a "plate" to represent the replication over k.

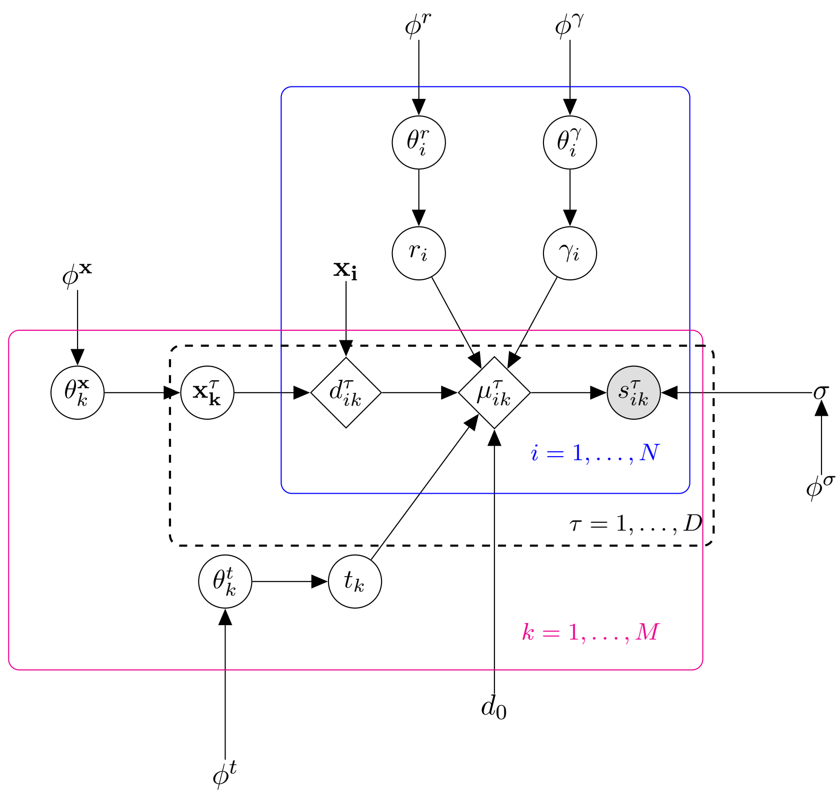

Fig. 2. Representation of the generative model as a graphical model. Open circles are random variables; the filled circles are observables. A diamond shape represents a quantity which is determined given the values of the variables feeding into it. The arrows express dependencies; a variable at the head of an arrow is dependent on the variable at its tail. Quantities not in shapes are constants. The surrounding boxes, or "plates" represent the replicated quantities. The blue plate shows the replications over the receivers; the magenta plate shows the replications over the transmitters; the dashed plate shows replication over multiple readings. At the periphery are hyperpriors (denoted as ϕs) which are sampled to determine the parameters of the priors (denoted as θs). Each replicated variable over i and k has its own prior, but the prior parameters are sampled from the same hyperprior, which cause the value of parameters which share a hyperprior to be similar.

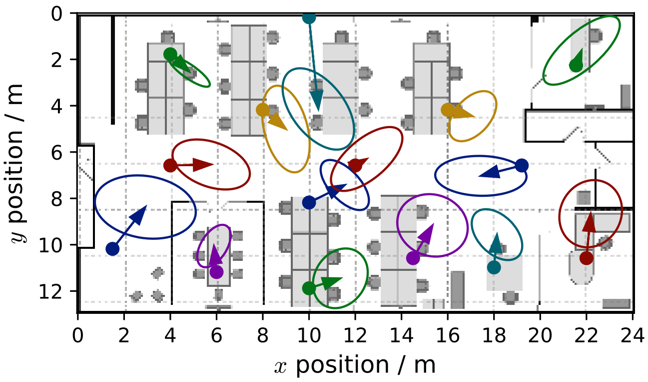

Fig. 3. Aggregated results for the position estimates. For each test location, an arrow points from the true position to the mean of the position estimates over all mobile devices and all time points. The ellipses represent the set of points of Mahalanobis distance 2.45 from the mean, as calculated from the empirical data covariance, which would enclose approximately 95% of the data if Gaussian distributed.

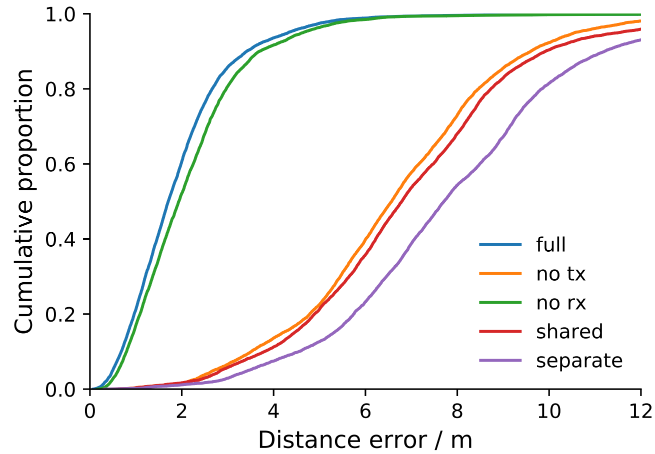

Fig. 4. Empirical CDFs of the distance error of the model variants, using the complete sensor set. In these experiments, treating the receivers as homogeneous affects performance only marginally, but including heterogeneity in transmitters improves performance significantly.

the median. Time bins were 10 seconds in duration, and a total of 7785 RSS vectors were collected.

The resulting data (known AP locations, AP RSS values, and mobile device RSS values) were entered into the model, and approximate Bayesian inference performed to obtain a posterior distribution over all unknown variables (see Section IV). These variables include, for example, the path loss exponents γ_i and missing RSS measurements, but for the purposes of evaluating the performance of the system we are mainly interested in the unknown device locations. Fig. 3 shows a summary of the estimated positions generated by the model. It can be seen that the resulting position estimates are consistent for any particular test position, having a precision of approximately $1\,\mathrm{m}$ to $2\,\mathrm{m}$, and accurate, with the average estimate error being $1.94\,\mathrm{m}$.

In order to determine the importance of explicitly modelling the heterogeneity of devices, we evaluated several model variants on the same position/RSS data, with different explanatory variables or a different structure. These models are labelled as: "full", the full model as described in Section III-A; "no tx", where the transmitters are treated as homogeneous, i.e. a model without the transmitter-dependent t_i parameters; "no rx", where the receivers are treated as homogeneous, i.e. a model without the receiver-dependent r_i and γ_i parameters; "separate", where receivers and transmitters are treated as individual and independent, i.e. a non-hierarchical model in which each AP has its own path-loss parameters; and "shared", a non-hierarchical model in which a single set of path-loss parameters are shared between all APs and without the t_i parameters.

As can be seen in Fig. 4, the "full" model, including all explanatory variables, is the best-performing model, having a superior distance error over the entire CDF. It is closely followed by the "no rx" model, with the other variants performing significantly worse, particularly the "separate" model. The "full" and "no rx" models were compared using the Widely Applicable Information Criterion (WAIC) [22], an estimate of the generalization error of the models, with calculated values of 366.4±0.2 and 386.8±0.2 per data point respectively. The WAIC is an estimate of the generalization error, so lower is better and it is on a logarithmic scale. This indicates that the "full" model should be strongly preferred.

As a final check on differences in the accuracy of the models, a t-test was carried out between the distance errors on the positions predicted by the two models, which indicated that the difference between them was statistically significant ($p = 1.4 \times 10^{-36}$).

Fig. 5 gives some insight into why including transmitter/receiver-dependent variables improves model performance. It shows a summary of the posterior means of the t_k and r_i parameters, grouped into different device types. First of all it is clear that the mobile devices have far weaker transmission power characteristics, with their median t being around $15\,\mathrm{dB}$ lower than that inferred for the APs. Including the transmission variables in the model allows it to adapt to this difference, which is important given that only inter-AP measurements have known distances for use in the inference procedure.

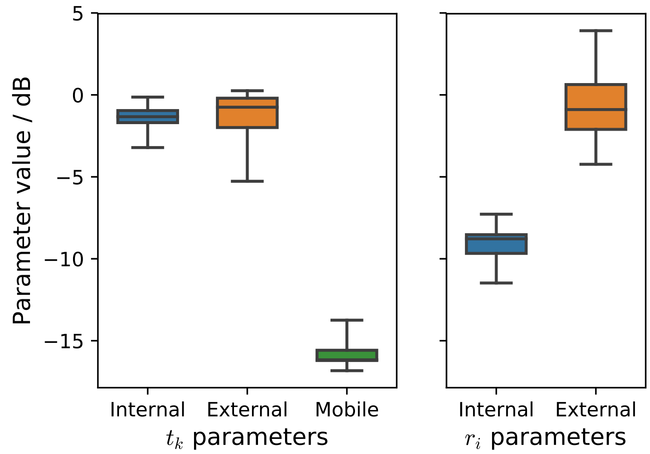

Fig. 5. Box plots of the posterior means of transmit and receive parameters, grouped by device type. On the left are the t_k parameters for the internal antenna APs, external antenna APs, and mobile devices, with medians of $-1.3\,\mathrm{dB}$, $-0.8\,\mathrm{dB}$ and $-16.2\,\mathrm{dB}$ respectively. On the right are the r_i parameters for the internal and external antenna APs, with medians of $-8.8\,\mathrm{dB}$ and $-0.9\,\mathrm{dB}$ respectively.

The figure also shows a similar plot for the receiver-dependent variables, in which a clear difference between the two AP sets can be seen: the internal antenna set has an average receive gain around $8\,\mathrm{dB}$ lower than the external antenna set. Not including this in the model would lead to large inaccuracies in the inferred distances from APs, which is reflected in the results shown in Figure 4. It should be noted that although the largest difference is between the sets of sensors, there is still significant within-set variation.

In order to determine whether the superior performance of the "full" model was solely due to heterogeneity in the APs, we evaluated the same set of model variants on two homogeneous sets of APs: those with external antennas, and those without. The results obtained are summarized in Table II, which shows the median and 95$^{\text{th}}$ percentile for all model variants and sensor sets. The same general pattern holds: the "full" model achieves the best results, with the removal of transmitter-dependent variables having a larger deleterious effect than the removal of receiver-dependent variables. This was confirmed by t-tests between the distance errors of the "full" model and the best alternative, which in all cases was the "shared" model: for the external antenna set, $p = 1.1 \times 10^{-5}$, and for the internal antenna set, $p = 5.8 \times 10^{-8}$. It should be noted, however, that the model variants without transmitter-dependent variables performed worse under the heterogeneous AP set, despite the larger number of RSS readings.

TABLE II
ACCURACY OF MODEL VARIANTS WITH DIFFERENT AP SETS.

AP set	Model variant	Median / m	95$^{\text{th}}$ percentile / m
All	Full	1.70	4.32
	No tx	6.57	10.56
	No rx	1.93	4.64
	Shared	6.79	11.52
	Separate	7.71	12.83
External antenna	Full	1.90	4.34
	No tx	3.01	6.18
	No rx	2.01	4.24
	Shared	3.72	7.51
	Separate	3.92	8.00
Internal antenna	Full	2.16	5.93
	No tx	3.52	8.02
	No rx	2.40	6.12
	Shared	3.72	8.26
	Separate	3.66	8.21

VI. CONCLUSION

We have introduced an indoor positioning system capable of localizing devices to a high degree of accuracy without the need for calibration or an extensive site-survey. The system is based on a Bayesian graphical model where unknowns are learned from inter-AP RSS measurements. We found that due to the fact that APs and mobile devices have considerable differences in their transmit/receive characteristics, a model with hierarchical structure explicitly modelling these differences significantly out-performed models lacking transmitter- and receiver-dependent variables. In particular, including a variable accounting for transmitter gain had a large effect, especially in the case of heterogeneity in the APs.

The use of a hierarchical Bayesian framework allows a simple way to introduce variations to components of the model. In this paper, we made use of this to account for the effect of varying transmission power of devices, but there are other potential variabilities which could be modelled, both in the environment and in the equipment. For instance, RSS values are strongly influenced by relative antenna orientation, which could be accounted for by including additional unknown state variables for devices to be located, along with position.

To progress this framework further, several directions could be pursued. First, as mentioned above, further differences could be investigated such as antenna type or device orientation. Second, further investigations in different environments, such as large, outdoor environments, or very dynamic environments, would help turn this into a widely usable tool. A third direction would be to introduce temporal structure into the position distribution, to increase accuracy of trajectory estimation. The model could then be interpreted as a state space model and operate in a similar fashion to Kalman or particle filtering.

Acknowledgements

We thank InnovateUK for supporting this work through the Knowledge Transfer Partnership KTP010049. We thank Mark Syme, James Wood, and Keith Orson for help setting up the experimental evaluation. Also thanks to Laura Dietz for making the tool [23] which we used to make Figures 1 and 2.

References

[1] K. Pahlavan, X. Li, and J. P. Mäkelä, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, 2002.

[2] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 13–32, 2009.

[3] R. Mautz and S. Tilch, "Survey of Optical Indoor Positioning Systems," in *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, 2011, pp. 1–7.

[4] S. A. R. Zekavat and R. M. Buehrer, Eds., *Handbook of Position Location*. John Wiley & Sons, 2012.

[5] P. Bahl and V. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings IEEE INFOCOM*, vol. 2, 2000, pp. 775–784.

[6] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005, pp. 205–218.

[7] A. K. M. M. Hossain and W. S. Soh, "A survey of calibration-free indoor positioning systems," *Computer Communications*, vol. 66, pp. 1–13, 2015.

[8] Y. Gwon and R. Jain, "Error characteristics and calibration-free techniques for wireless lan-based location estimation," in *Proceedings of the second international workshop on Mobility management & wireless access protocols*. ACM, 2004, pp. 2–9.

[9] H. Lim, L. C. Kung, J. C. Hou, and H. Luo, "Zero-configuration, robust indoor localization: Theory and experimentation," in *Proceedings IEEE INFOCOM*, 2006, pp. 1–13.

[10] L. F. M. de Moraes and B. A. A. Nunes, "Calibration-free WLAN location system based on dynamic mapping of signal strength," *Proceedings of the international workshop on Mobility management and wireless access - MobiWac '06*, pp. 92–99, 2006.

[11] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 2010, pp. 173–184.

[12] D. Madigan, E. Einahrawy, R. P. Martin, W. H. Ju, P. Krishnan, and a. S. Krishnakumar, "Bayesian indoor positioning systems," in *IEEE Conference on Computer and Communications*, vol. 2, 2005, pp. 1217–1227.

[13] S. H. Fang, C. H. Wang, S. M. Chiou, and P. Lin, "Calibration-free approaches for robust Wi-Fi positioning against device diversity: A performance comparison," in *IEEE Vehicular Technology Conference*, 2012.

[14] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, 2004, pp. 70–84.

[15] A. W. Tsui, Y. H. Chuang, and H. H. Chu, "Unsupervised learning for solving RSS hardware variance problem in WiFi localization," *Mobile Networks and Applications*, vol. 14, no. 5, pp. 677–691, 2009.

[16] A. K. M. M. Hossain, H. N. Van, Y. Jin, and W.-S. Soh, "Indoor Localization Using Multiple Wireless Technologies," in *2007 IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems*, 2007, pp. 1–8.

[17] A. Goswami, L. E. Ortiz, and S. R. Das, "WiGEM : A Learning-Based Approach for Indoor Localization," in *Proceedings of the Seventh Conference on emerging Networking EXperiments and Technologies*. ACM, 2011.

[18] W. L. Buntine, "Operations for learning with graphical models," *Journal of Artificial Intelligence Research*, vol. 2, pp. 159–225, 1994.

[19] A. Gelman, J. G. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis*, 3rd ed. CRC Press, 2013.

[20] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3," *PeerJ Computer Science*, vol. 2, p. e55, 2016.

[21] R. M. Neal, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press, 2011, ch. 5 — MCMC using Hamiltonian dynamics, available on arXiv: https://arxiv.org/pdf/1206.1901.pdf.

[22] S. Watanabe, "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, vol. 11, pp. 3571–3594, 2010.

[23] L. Dietz, "Directed factor graph notation for generative models," University of New Hampshire, Tech. Rep., 2010, available from https://github.com/jluttine/tikz-bayesnet.