Occupancy Estimation with Wireless Monitoring Devices and Application-Specific Antennas

Pranay Eedara, Hai Li, Nagaraj Janakiraman, N. R. Anudeep Tungala,
Jean-Francois Chamberland, Gregory H. Huff
Department of Electrical and Computer Engineering, Texas A&M University

***Abstract*—The rapid proliferation of wireless devices offers new means to infer current conditions surrounding mobile users, their locations, and their actions. This is especially true for industrial, scientific and medical (ISM) radio bands where communication protocols are often open and information about neighboring devices abounds. This article examines efficient occupancy estimation based on Wi-Fi metadata, with an emphasis on algorithms attuned to directional antenna technology. Certain Wi-Fi interfaces can be switched into monitoring mode, an operating state where all local packets are observed and recorded. Using a network of such monitoring sensors, it is then possible to estimate the number of active devices within a specific area. The envisioned estimators take as input received signal strength indicators and media access control addresses. By using directional monitoring antennas, one can gain additional and/or more discriminating information about current conditions, thereby yielding enhanced occupancy estimates. This work introduces novel estimation algorithms and characterizes the performance gains associated with RF-aware sensing devices. Experimental results based on a prototype implementation of this distributed monitoring system provide further supporting evidence for the proposed techniques.**

***Index Terms*—Occupancy estimation, maximum likelihood, directional antennas, statistical information processing, communication systems.**

I. Introduction

The advent of smartphones has changed the wireless landscape considerably. A growing mobile data traffic worldwide is forcing service providers and users to leverage various modes of connectivity. Cisco Systems predicts in their Visual Network Index that 55 percent of total mobile data traffic will be offloaded onto fixed networks through Wi-Fi access points and femtocells by 2020 [1]. This translates into several exabytes of mobile data passing through local area wireless networks. A noteworthy aspect of Wi-Fi technology is the fact that transmitted data packets can be captured passively and subsequently analyzed in a straightforward fashion. The existence of application programming interfaces for monitoring wireless traffic like `pcap` and `WinPcap`, and the availability of fully-featured programs such as Wireshark and Kismet make traffic analysis a straightforward task. This fact, combined with a very large smartphone penetration rate and recurrent exchanges of Wi-Fi data packets between smartphones and access points, offers novel and unique means to infer properties of the physical world [2].

This situation has piqued the interest of researchers and hobbyists alike, leading to a vast and growing body of literature on inference tasks based on wireless technology [3], [4], [5], [6], [7]. The wardriving movement illustrates well how Wi-Fi signals can be tied to physical locations. Radio frequency (RF) signals have been employed as a basis for self-localization, source localization, and the more intricate problem of nodes identifying the structure of their own ad hoc wireless network [8], [9], [10], [11], [12], [13], [14]; with applications ranging from robotics to location-based advertising. In this article, we are interested in the related problem of estimating room occupancy based on Wi-Fi activity. Occupancy estimation based on wireless signals is the focus of recent work [15], [16]. The estimates can be employed in the contexts of building code compliance, smart homes with HVAC management, and the assessment of emergency egress from buildings [17], [18].

The ability of cyber-physical systems to conduct such inference tasks is closely linked to their intrinsic characteristics. Ongoing efforts in metrology illustrate well how the design of sensing devices is connected to specific application scenarios [19], [20]. In a similar fashion, the radiation patterns of sensing antennas can play a determinant role in parameter estimation based on RF monitoring. This fact is well understood in the antennas and propagation community, as it underlies much of the research work in antenna design for radar, communication, and other applications. However, the domain knowledge related to electromagnetics is yet to be fully incorporated in the design and analysis of inference using wireless systems. Indeed, RF antennas are often perceived as a commodity and the potential impacts of their operating characteristics are frequently overlooked in the communication and signal processing literature.

While most consumer-based RF products are aimed at wireless communication, it is possible to design or carefully select antennas specifically for sensing and monitoring. Furthermore, with the emergence of reconfigurable antenna technology, a monitoring system can dynamically adapt to changing environmental conditions or an evolving system objective. The goal of this article is to revisit occupancy estimation, albeit giving due consideration to the radiation characteristics of the sensing devices. Specifically, we are interested in estimating the number of active wireless agents located within a prescribed area using RF monitoring through multiple distributed agents.

This material is based upon work supported, in part, by the National Science Foundation (NSF) under Grant No. CCF-1619085. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

In the envisioned setting, the data acquired by these various devices is aggregated over the Internet, and the inference task is performed at a central location.

While producing occupancy estimates, the antenna radiation patterns of the different agents can be integrated into the inference task. Our analysis framework enables the study of potential performance gains associated with the careful analysis of RF signals. It also offers a novel perspective on antenna selection for RF monitoring in the context of occupancy estimation. Finally, it provides a means to better understand antenna designs for inference tasks, and presents a unique opportunity to study how reconfigurable antennas can enhance the performance of monitoring systems. Altogether, the article gives a new perspective on the interplay between antenna design and RF monitoring over wireless networks with distributed agents.

The remainder of this article is structured as follows. The problem formulation is described in Section II. Three estimation schemes are introduced in Section III. Performance results corresponding to the proposed schemes for sensing systems with isotropic and directional antennas appear in Section IV. Experimental findings derived from a prototype implementation are contained in Section V, along with a brief description of our experimental setup. Insights, guidelines and final remarks conclude this article.

II. Problem Formulation

Consider a prescribed neighborhood of interest with a set geometric shape. For simplicity, we take this region to be rectangular. Several RF monitoring devices are located about this area, in an arbitrary fashion. Each monitoring device is aware of its own location and orientation, and it knows the radiation pattern of its antenna. In the proposed framework, the monitoring devices are connected to the Internet through a wireline infrastructure, and they are relaying sampled data to a central entity where data is aggregated for analysis. In addition, several wireless agents are randomly located within this environment. Each agent can lie inside or outside an area of interest. We use $\mathcal{A}_{\mathrm{t}}$ to denote the target area and $\mathcal{A}_{\mathrm{o}}$ to indicate its complement. All the wireless agents periodically transmit data packets, thereby revealing partial information about their existence and whereabouts to monitoring devices. Every agent uses a unique MAC address and, consequently, the packets associated with separate devices can be differentiated. Thus, using MAC addresses, it is possible to count the total number of active agents recorded by monitoring devices, although their precise locations remain unknown. A notional diagram of the problem formulation appears in Fig. 1.

In this study, we assume that the wireless users are quasi-static and an inference task is performed over a single monitoring period. Although it is possible to incorporate moving agents and streaming observations into the problem formulation, these features render the analysis more challenging and they are not necessary to achieve our goal of better understanding the interplay between antenna profiles and RF monitoring. As such, advanced mobility models and their ramifications are relegated to future work in favor of a simpler,

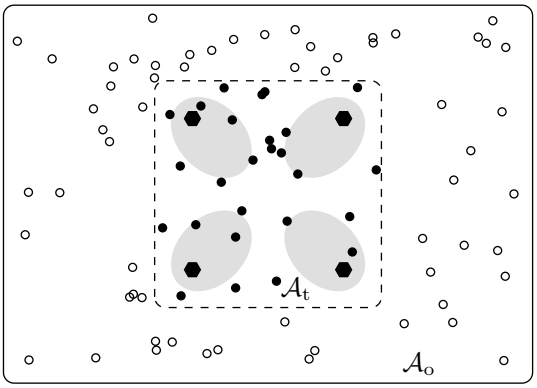

Fig. 1. This notional diagram illustrates the inference task, with the periphery of target area delineated by the dashed line. Hexagons denote the locations of monitoring devices equipped with directional antennas. Agents within the zone of interest are in black, whereas outside agents appear as circles. The objective is to estimate occupancy within the target area.

explicative model suited to our main purpose. For convenience, the random locations of the wireless devices are aggregated into a single vector, $\underline{\mathbf{U}} = (\mathbf{U}_1, \ldots, \mathbf{U}_{n_{\mathrm{a}}})$. Wireless agents are equipped with vertically polarized, isotropic antennas. The power radiated by one agent is therefore uniform in all directions on the plane. The transmit power for every mobile agent is known and equal to the limit prescribed by the Federal Communications Commission (FCC) for the industrial, scientific and medical (ISM) radio bands.

For the purpose of analysis, we assume that the energy captured by a monitoring device comes primarily from a line-of-sight path. Over short distances, the signal strength of the electromagnetic wave coming from this agent is governed by the free-space path loss. The received signal strength from source j to sensing unit i can be expressed as

$$P_{ij}[\mathrm{dBm}] = A + B\log_{10}(d_{ij}) + L_{ij} + G_i(\phi_{ij}) \qquad (1)$$

where A and B are the mean decay parameters, d_{ij} represents the Euclidean distance between the source and the measurement device, L_{ij} denotes shadow fading, and $G_i(\cdot)$ is the antenna gain of the sensing unit. Explicitly, suppose that the sensing agent is located at point $\mathbf{s}_i = (s_{1i}, s_{2i})$ and the signal originates from $\mathbf{u}_j = (u_{1j}, u_{2j})$; then, the distance between the two points is equal to

$$\begin{aligned} d_{ij} &= d(\mathbf{s}_i, \mathbf{u}_j) = \|\mathbf{u}_j - \mathbf{s}_i\|_2 \\ &= \sqrt{(u_{1j} - s_{1i})^2 + (u_{2j} - s_{2i})^2}. \end{aligned}$$

Likewise, the angle of incidence of the electromagnetic wave onto the sensing unit is given by

$$\phi_{ij} = \angle(\mathbf{s}_i, \mathbf{u}_j) = \mathrm{atan2}(u_{2j} - s_{2i}, u_{1j} - s_{1i}),$$

where $\mathrm{atan2}(\cdot,\cdot)$ is the two-argument variant of the arctangent function. The antenna gain $G_i(\cdot)$[dB] may depend on the antenna characteristics of device i, its current orientation, the angle of incidence of the incoming signal, and its polarization.

The shadow fading components $\{L_{ij}\}$ are assumed to form a set of independent and identically distributed random variables, each with a log-normal distribution. In the logarithmic

domain, the probability density function common to all fading components, $\{L_{ij}\}$, is

$$f_{L_{ij}}(\ell) = \frac{1}{\sqrt{2\pi}\sigma_{\rm s}} \exp\left(-\frac{\ell^2}{2\sigma_{\rm s}^2}\right) \tag{2}$$

where $\sigma_{\rm s}$ is the logarithmic standard deviation of the shadow fading, expressed in natural units.

The observed information available at the fusion center for the purpose of inference takes the form of a sequence of vectors $\underline{\mathbf{P}} = (\mathbf{P}_1, \ldots, \mathbf{P}_{n_{\rm a}})$, with each information vector corresponding to a particular mobile agent. The length of vector element $\mathbf{P}_j$ corresponds to the total number of sensing devices monitoring the field of interest. Its entries are the signal strengths captured by the sensors, $\mathbf{P}_j = (P_{1j}, \ldots, P_{n_{\rm s}j})$ where $n_{\rm s}$ is the number of monitoring devices. The fusion center possesses side information in the form of the locations of sensing devices, their orientations and antenna properties.

The number and locations of the wireless units contained within the region of interest is assumed to form a Poisson point process with intensity $\lambda_{\rm t}$. If we denote the number of active devices that lies therein by $R_{\rm t}$, we get

$$\Pr(R_{\rm t} = r_{\rm t}) = \frac{(\lambda_{\rm t} A_{\rm t})^{r_{\rm t}}}{r_{\rm t}!} e^{-A_{\rm t}\lambda_{\rm t}} \quad r_{\rm t} = 0, 1, \ldots$$

where $A_{\rm t}$ represents the area of the target region $\mathcal{A}_{\rm t}$. Similarly, the device count and their locations outside of the zone of interest is taken to be a Poisson point process with intensity $\lambda_{\rm o}$. The distribution of $R_{\rm o}$, the number of devices outside the target region is

$$\Pr(R_{\rm o} = r_{\rm o}) = \frac{(\lambda_{\rm o} A_{\rm o})^{r_{\rm o}}}{r_{\rm o}!} e^{-A_{\rm o}\lambda_{\rm o}} \quad r_{\rm o} = 0, 1, \ldots$$

where $A_{\rm o}$ is the area of the surveyed neighborhood, excluding the target region $\mathcal{A}_{\rm t}$. Note that $R_{\rm t}$ and $R_{\rm o}$ can be determined from the locations of the agents and, as such, it is sometimes useful to think of them as functions of $\underline{\mathbf{U}}$. The inference tasks at hand consist in estimating model parameters and occupancy based on the information collected by the monitoring sensors.

III. Estimation Schemes

In this section, we introduce estimation schemes tailored to different application scenarios. The suitability of a particular estimator depends on the amount of side information available at the fusion center. In situations where the monitoring system is employed repetitively to estimate the occupancy of a prescribed area, prior parameters can be formed and the Bayesian framework applies. On the other hand, if the system is deployed in a new environment and used once, a classical formulation may be more appropriate. This will become manifest shortly. The development of these estimation schemes necessitates extra notation. Whenever appropriate, we use uppercase letters for random variables and lowercase letters for realizations or arguments.

A. Bayes Estimation of Occupancy

The Bayes estimation framework assumes that parameters $\lambda_{\rm t}$ and $\lambda_{\rm o}$ are known. As such, the corresponding Poisson distributions act as prior probabilities for the numbers of units within and outside the target area, respectively. Our goal is to estimate the number of active units located within the target area based on observed data $\underline{\mathbf{P}}$. A first step in deriving this estimator is to get the posterior distribution of $R_{\rm t}$, conditional on the gathered data,

$$\begin{aligned}\Pr\left(R_{\rm t} = r_{\rm t} | \underline{\mathbf{P}} = \underline{\mathbf{p}}\right) &= \int_{\underline{\mathbf{u}}} \mathbb{1}_{\{R_{\rm t}(\underline{\mathbf{u}}) = r_{\rm t}\}} f_{\underline{\mathbf{U}}|\underline{\mathbf{P}}}\left(\underline{\mathbf{u}}|\underline{\mathbf{p}}\right) d\underline{\mathbf{u}} \\ &= \int_{\{\underline{\mathbf{u}}: R_{\rm t}(\underline{\mathbf{u}}) = r_{\rm t}, R_{\rm o}(\underline{\mathbf{u}}) = r_{\rm o}\}} f_{\underline{\mathbf{U}}|\underline{\mathbf{P}}}\left(\underline{\mathbf{u}}|\underline{\mathbf{p}}\right) d\underline{\mathbf{u}} \\ &= \int_{\{\underline{\mathbf{u}}: R_{\rm t}(\underline{\mathbf{u}}) = r_{\rm t}, R_{\rm o}(\underline{\mathbf{u}}) = r_{\rm o}\}} \frac{f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}})}{f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right)} d\underline{\mathbf{u}}\end{aligned} \tag{3}$$

where $\underline{\mathbf{U}}$ represents the random positions of the wireless units and $\mathbb{1}_{\{\cdot\}}$ is a shorthand notation for the standard indicator function. The first equality in (3) is based on the fact that the probability of an event can be expressed as the expectation of an indicator function.

The collection of device locations contains much information about the problem at hand. For instance, the size of $\underline{\mathbf{U}}$ is equal to the number of active wireless agents. Moreover, it is possible to extract from $\underline{\mathbf{U}}$ the numbers of units within the target area $R_{\rm t}\left(\underline{\mathbf{U}}\right)$, and the number of units outside the target areas $R_{\rm o}\left(\underline{\mathbf{U}}\right)$. These implicit relations delineate the regions over which the multi-dimensional integrals in (3) are taken.

Because the two Poisson point processes are independent, the distribution of $\underline{\mathbf{U}}$ can be written as

$$\begin{aligned} f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}) &= \frac{1}{A_{\rm t}^{R_{\rm t}(\underline{\mathbf{u}})}} \frac{(\lambda_{\rm t} A_{\rm t})^{R_{\rm t}(\underline{\mathbf{u}})}}{(R_{\rm t}(\underline{\mathbf{u}}))!} e^{-A_{\rm t}\lambda_{\rm t}} \\ &\quad \times \frac{1}{A_{\rm o}^{R_{\rm o}(\underline{\mathbf{u}})}} \frac{(\lambda_{\rm o} A_{\rm o})^{R_{\rm o}(\underline{\mathbf{u}})}}{(R_{\rm o}(\underline{\mathbf{u}}))!} e^{-A_{\rm o}\lambda_{\rm o}} \\ &= \frac{\lambda_{\rm t}^{R_{\rm t}(\underline{\mathbf{u}})}}{(R_{\rm t}(\underline{\mathbf{u}}))!} \frac{\lambda_{\rm o}^{R_{\rm o}(\underline{\mathbf{u}})}}{(R_{\rm o}(\underline{\mathbf{u}}))!} e^{-A_{\rm t}\lambda_{\rm t} - A_{\rm o}\lambda_{\rm o}}. \end{aligned} \tag{4}$$

The distribution of the received power from an agent standing at a specific location is obtained based on the description contained in Section II. For instance, the distribution of the power vector $\mathbf{P}_j$ associated with agent j, conditional on unit location $\mathbf{U}_j = \mathbf{u}_j$, is equal to

$$\begin{aligned} &f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) \\ &= \prod_{i=1}^{n_{\rm s}} f_{L_{ij}}(p_{ij} - A - B\log_{10}(d_{ij}) - G_i(\phi_{ij})) \\ &= \frac{1}{(2\pi\sigma_{\rm s}^2)^{\frac{n_{\rm s}}{2}}} \prod_{i=1}^{n_{\rm s}} e^{-\frac{(p_{ij} - A - B\log_{10}(d_{ij}) - G_i(\phi_{ij}))^2}{2\sigma_{\rm s}^2}} \\ &= \left(2\pi\sigma_{\rm s}^2\right)^{-\frac{n_{\rm s}}{2}} e^{-\frac{\sum_{i=1}^{n_{\rm s}}(p_{ij} - A - B\log_{10}(d_{ij}) - G_i(\phi_{ij}))^2}{2\sigma_{\rm s}^2}}. \end{aligned} \tag{5}$$

The conditional distribution for the gathered data $\underline{\mathbf{P}}$, given $\underline{\mathbf{U}} = \underline{\mathbf{u}}$, is the product of the marginal distributions

$$f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) = \prod_{j=1}^{n_{\rm a}} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j).$$

At this point, we can revisit (3), the conditional distribution of $R_{\rm t}$ given the received signal strength readings. Substituting

the marginal distribution of $\underline{\mathbf{U}}$ introduced in (4) and the product form for the conditional distribution of $\underline{\mathbf{P}}$ given $\underline{\mathbf{U}}$ shown above, we get

$$\begin{aligned}
&\Pr\left(R_{\mathrm{t}}=r_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}}\right)\\
&\quad=\int_{\{\underline{\mathbf{u}}:R_{\mathrm{t}}(\underline{\mathbf{u}})=r_{\mathrm{t}},R_{\mathrm{o}}(\underline{\mathbf{u}})=r_{\mathrm{o}}\}}\frac{f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right)f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}})}{f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right)}d\underline{\mathbf{u}}\\
&\quad=\int_{\{\underline{\mathbf{u}}:R_{\mathrm{t}}(\underline{\mathbf{u}})=r_{\mathrm{t}},R_{\mathrm{o}}(\underline{\mathbf{u}})=r_{\mathrm{o}}\}}\prod_{j=1}^{n_{\mathrm{a}}}\frac{f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)}{f_{\underline{\mathbf{P}}}(\underline{\mathbf{p}})}\\
&\qquad\times\frac{\lambda_{\mathrm{t}}^{R_{\mathrm{t}}(\underline{\mathbf{u}})}}{(R_{\mathrm{t}}(\underline{\mathbf{u}}))!}\frac{\lambda_{\mathrm{o}}^{R_{\mathrm{o}}(\underline{\mathbf{u}})}}{(R_{\mathrm{o}}(\underline{\mathbf{u}}))!}e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}-A_{\mathrm{o}}\lambda_{\mathrm{o}}}d\underline{\mathbf{u}}.
\end{aligned}$$

Introducing a convenient notation for pertinent integrals, we obtain the desired form for the posterior distribution of R_{t},

$$\begin{aligned}
&\Pr\left(R_{\mathrm{t}}=r_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}}\right)\\
&\quad=\sum_{\{\mathbb{I}\subset[n_{\mathrm{a}}]:|\mathbb{I}|=r_{\mathrm{t}}\}}\frac{\lambda_{\mathrm{t}}^{r_{\mathrm{t}}}\lambda_{\mathrm{o}}^{r_{\mathrm{o}}}e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}-A_{\mathrm{o}}\lambda_{\mathrm{o}}}}{r_{\mathrm{t}}!r_{\mathrm{o}}!f_{\underline{\mathbf{P}}}(\underline{\mathbf{p}})}\\
&\qquad\times\prod_{j\in\mathbb{I}}\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)\prod_{j\in\mathbb{I}^{\mathrm{c}}}\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)
\end{aligned}\tag{6}$$

where the integral components are defined by

$$\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)=\int_{\mathcal{A}_{\mathrm{t}}}f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)d\mathbf{u}_j\tag{7}$$

$$\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)=\int_{\mathcal{A}_{\mathrm{o}}}f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)d\mathbf{u}_j.\tag{8}$$

Throughout, we use the popular notation $[n_{\mathrm{a}}]$ to denote $\{1,\ldots,n_{\mathrm{a}}\}$, the set of natural numbers from one to n_{a}.

Equation (6) offers an algorithmic blueprint to compute the conditional distribution of R_{t}. The integrals in (7) and (8) are taken over structured two-dimensional sets. They pose no particular computational challenge, and the numerical evaluations can be parallelized to take advantage of modern computer architectures. Furthermore, the calculated values can be leveraged to compute the marginal distribution of $\underline{\mathbf{P}}$,

$$\begin{aligned}
&f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right)=\int_{\{\underline{\mathbf{u}}:R_{\mathrm{t}}(\underline{\mathbf{u}})+R_{\mathrm{o}}(\underline{\mathbf{u}})=n_{\mathrm{a}}\}}f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right)f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}})d\underline{\mathbf{u}}\\
&=\sum_{(r_{\mathrm{t}},r_{\mathrm{o}}):r_{\mathrm{t}}+r_{\mathrm{o}}=n_{\mathrm{a}}}\ \sum_{\{\mathbb{I}\subset[n_{\mathrm{a}}]:|\mathbb{I}|=r_{\mathrm{t}}\}}\frac{\lambda_{\mathrm{t}}^{r_{\mathrm{t}}}\lambda_{\mathrm{o}}^{r_{\mathrm{o}}}}{r_{\mathrm{t}}!r_{\mathrm{o}}!}e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}-A_{\mathrm{o}}\lambda_{\mathrm{o}}}\\
&\qquad\times\prod_{j\in\mathbb{I}}\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)\prod_{j\in\mathbb{I}^{\mathrm{c}}}\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j).
\end{aligned}\tag{9}$$

As mentioned above, evaluating the collection of integrals $\{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(\cdot),\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(\cdot)\}$ is not a particularly difficult task, and its computational complexity grows linearly with n_{a}. Computing the conditional distribution of R_{t} given $\underline{\mathbf{P}}$, which appears in (6), and the closely related quantity $f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right)$ may appear more difficult, as it entails taking sums over subsets of $[n_{\mathrm{a}}]$. However, while the cardinality of the power set of $[n_{\mathrm{a}}]$ grows exponentially fast, the posterior distribution of R_{t} can be computed efficiently using generating functions [21]. Thus, for practical values of n_{a}, getting the conditional distribution is manageable.

Once the posterior distribution of R_{t} conditional upon $\underline{\mathbf{P}}$ is obtained, its mean can be employed as an estimate,

$$\begin{aligned}
\hat{R}_{\mathrm{t}}\left(\underline{\mathbf{p}}\right)&=\mathrm{E}\left[R_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}}\right]\\
&=\sum_{r_{\mathrm{t}}=0}^{n_{\mathrm{a}}}r_{\mathrm{t}}\Pr\left(R_{\mathrm{t}}=r_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}}\right).
\end{aligned}\tag{10}$$

To characterize the performance of Bayes estimators, we adopt the Bayesian mean squared error (BMSE),

$$\mathrm{BMSE}\left[\hat{R}_{\mathrm{t}}\right]=\mathrm{E}\left[\left(\hat{R}_{\mathrm{t}}\left(\underline{\mathbf{P}}\right)-R_{\mathrm{t}}\right)^2\right].\tag{11}$$

Numerically, the BMSE can be approximated through empirical averaging over a large sample,

$$\mathrm{BMSE}\left[\hat{R}_{\mathrm{t}}\right]\approx\frac{1}{M}\sum_{m=1}^{M}\left(\hat{R}_{\mathrm{t}}^{(m)}\left(\underline{\mathbf{P}}^{(m)}\right)-R_{\mathrm{t}}^{(m)}\right)^2.\tag{12}$$

It is well known that the mean of the posterior distribution minimizes the BMSE [22], [23], [24]. In this framework, the optimal estimator is thus given by (10). Based on this result, we can evaluate the performance of our estimator for different placements of the sensing devices $\underline{\mathbf{s}}=(\mathbf{s}_1,\ldots,\mathbf{s}_{n_{\mathrm{s}}})$, various antenna radiation patterns, and device orientations. These characteristics implicitly shape the conditional distributions $\left\{f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)\right\}$, and they consequently impact overall system performance.

B. Estimating Intensity Parameters

In the previous section, we assume that intensity parameters λ_{t} and λ_{o} are available. Still, there exist scenarios where these parameters are not known a priori. In such cases, it is possible to produce estimates for these quantities based on the observed data. This task forms the primary focus of this section. To estimate the values of the intensity parameters, we embrace the classical approach and adopt maximum-likelihood estimation as a basis for statistical inference [24].

Again, let $\underline{\mathbf{U}}$ represent the positions of all the wireless units. As before, R_{t} and R_{o} can be obtained as functions of $\underline{\mathbf{U}}$, and the size of $\underline{\mathbf{U}}$ varies depending on the realization of the process. The distribution of (4) still applies; yet to conform with the classical statistical framework, we make the dependence on the intensity parameters explicit,

$$f_{\underline{\mathbf{U}}}\left(\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right)=\frac{\lambda_{\mathrm{t}}^{R_{\mathrm{t}}(\underline{\mathbf{u}})}}{(R_{\mathrm{t}}(\underline{\mathbf{u}}))!}\frac{\lambda_{\mathrm{o}}^{R_{\mathrm{o}}(\underline{\mathbf{u}})}}{(R_{\mathrm{o}}(\underline{\mathbf{u}}))!}e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}-A_{\mathrm{o}}\lambda_{\mathrm{o}}}.$$

In the current model, $\underline{\mathbf{P}}$ forms the observed data, $\underline{\mathbf{U}}$ is composed of unobserved latent variables, and parameters λ_{t} and λ_{o} are the unknowns. The likelihood function for the latter parameters is equal to

$$\begin{aligned}
\mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}},\underline{\mathbf{u}}\right)&=f_{\underline{\mathbf{P}},\underline{\mathbf{U}}}\left(\underline{\mathbf{p}},\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right)\\
&=f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right)f_{\underline{\mathbf{U}}}\left(\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right).
\end{aligned}$$

This, in turn, leads to the marginal likelihood of the observed data found below,

$$\begin{aligned}
&\mathcal{L}\left(\lambda_{\mathrm{t}}, \lambda_{\mathrm{o}}; \underline{\mathbf{p}}\right) \\
&= \int_{\{\underline{\mathbf{u}}: R_{\mathrm{t}}(\underline{\mathbf{u}}) + R_{\mathrm{o}}(\underline{\mathbf{u}}) = n_{\mathrm{a}}\}} f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}\left(\underline{\mathbf{u}}; \lambda_{\mathrm{t}}, \lambda_{\mathrm{o}}\right) d\underline{\mathbf{u}} \\
&= e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}} - A_{\mathrm{o}}\lambda_{\mathrm{o}}} \sum_{(r_{\mathrm{t}}, r_{\mathrm{o}}): r_{\mathrm{t}} + r_{\mathrm{o}} = n_{\mathrm{a}}} \frac{\lambda_{\mathrm{t}}^{r_{\mathrm{t}}} \lambda_{\mathrm{o}}^{r_{\mathrm{o}}}}{r_{\mathrm{t}}! r_{\mathrm{o}}!} \\
&\qquad \times \sum_{\{\mathbb{I} \subset [n_{\mathrm{a}}]: |\mathbb{I}| = r_{\mathrm{t}}\}} \prod_{j \in \mathbb{I}} \mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j) \prod_{j \in \mathbb{I}^{\mathrm{c}}} \mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j).
\end{aligned} \tag{13}$$

The marginalization is performed through integration. Explicit forms for the integral components $\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)$ and $\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)$ appear in (7) and (8), respectively. Again, the high-dimensional integral breaks down into products of two-dimensional integrals. This renders the computational problem tractable. The maximum-likelihood estimate for a set of observations $\underline{\mathbf{p}}$ can be obtained by performing a two-dimensional optimization on the objective function in (13). Fortunately, this two-dimensional optimization task reduces to a one-dimensional problem. We present this property formally in the proposition below.

Proposition 1: The likelihood function of the observed data possesses the following property,

$$\max_{\lambda_{\mathrm{t}}, \lambda_{\mathrm{o}}} \mathcal{L}\left(\lambda_{\mathrm{t}}, \lambda_{\mathrm{o}}; \underline{\mathbf{p}}\right) = \max_{\alpha} \mathcal{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha, \frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha); \underline{\mathbf{p}}\right). \tag{14}$$

That is, the two-dimensional optimization simplifies to a maximization over one variable.

Proof: See Appendix A. ∎

In view of Proposition 1, the maximum-likelihood estimator for $(\lambda_{\mathrm{t}}, \lambda_{\mathrm{o}})$ can be implemented by first solving the lower-dimensional optimization problem and then extracting values for parameters λ_{t} and λ_{o}. To further analyze this task, we expand the objective function in terms of α,

$$\begin{aligned}
&\mathcal{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha, \frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha); \underline{\mathbf{p}}\right) \\
&= \sum_{(r_{\mathrm{t}}, r_{\mathrm{o}}): r_{\mathrm{t}} + r_{\mathrm{o}} = n_{\mathrm{a}}} \frac{e^{-n_{\mathrm{a}}} n_{\mathrm{a}}^{n_{\mathrm{a}}}}{r_{\mathrm{t}}! r_{\mathrm{o}}!} \left(\frac{\alpha}{A_{\mathrm{t}}}\right)^{r_{\mathrm{t}}} \left(\frac{1-\alpha}{A_{\mathrm{o}}}\right)^{r_{\mathrm{o}}} \\
&\qquad \times \sum_{\{\mathbb{I} \subset [n_{\mathrm{a}}]: |\mathbb{I}| = r_{\mathrm{t}}\}} \prod_{j \in \mathbb{I}} \mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j) \prod_{j \in \mathbb{I}^{\mathrm{c}}} \mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)
\end{aligned}$$

This objective function is a polynomial in a single indeterminate, and the maximization takes place over the compact interval $\alpha \in [0, 1]$. Although this problem does not appear to admit a closed-form expression, obtaining a solution through numerical methods is possible for systems with a realistic number of active devices. One approach is to first compute the derivative of the likelihood function, and then employ standard means to identify the roots of the resulting polynomial. These roots correspond to critical points of the likelihood function. All the roots that fall within the interval $[0, 1]$, along with the boundary points $\alpha \in \{0, 1\}$, can then be tested as candidate maximizers. For typical values of parameters λ_{t} and λ_{o}, this approach offers a pragmatic solution to the estimation problem. In this setting, the mean squared error associated with the estimation procedure becomes

$$\begin{aligned}
&\mathrm{MSE}\left[\hat{\lambda}_{\mathrm{t}}\left(\underline{\mathbf{P}}\right), \hat{\lambda}_{\mathrm{o}}\left(\underline{\mathbf{P}}\right)\right] \\
&= \mathrm{E}_{\lambda_{\mathrm{t}}, \lambda_{\mathrm{o}}}\left[\left(\hat{\lambda}_{\mathrm{t}}\left(\underline{\mathbf{P}}\right) - \lambda_{\mathrm{t}}\right)^2 + \left(\hat{\lambda}_{\mathrm{o}}\left(\underline{\mathbf{P}}\right) - \lambda_{\mathrm{o}}\right)^2\right].
\end{aligned}$$

After estimating the intensity parameters, a natural approach to assessing the number of active units inside the target region is to employ the technique discussed in Section III-A, albeit with intensity estimates $\hat{\lambda}_{\mathrm{t}}$ and $\hat{\lambda}_{\mathrm{o}}$;

$$\begin{aligned}
\hat{R}_{\mathrm{t}}\left(\underline{\mathbf{p}}\right) &= \mathrm{E}_{\hat{\lambda}_{\mathrm{t}}, \hat{\lambda}_{\mathrm{o}}}\left[R_{\mathrm{t}} | \underline{\mathbf{P}} = \underline{\mathbf{p}}\right] \\
&= \sum_{r_{\mathrm{t}}=0}^{n_{\mathrm{a}}} r_{\mathrm{t}} \Pr\left(R_{\mathrm{t}} = r_{\mathrm{t}} | \underline{\mathbf{P}} = \underline{\mathbf{p}}; \hat{\lambda}_{\mathrm{t}}, \hat{\lambda}_{\mathrm{o}}\right)
\end{aligned} \tag{15}$$

where the probability of R_{t} conditional on $\underline{\mathbf{P}} = \underline{\mathbf{p}}$ is taken under model parameters $\hat{\lambda}_{\mathrm{t}}$ and $\hat{\lambda}_{\mathrm{o}}$.

C. Iterative Algorithm

The techniques introduced above constitute powerful tools to estimate room occupancy and model parameters. Yet, when the total number of active agents is large, these techniques can become computationally challenging. To circumvent this potential difficulty, we devise an iterative estimation strategy. To begin, we emphasize that under the current model, a stochastically equivalent way to obtain a realization is to first generate the total number of active devices according to Poisson distribution

$$\Pr(N_{\mathrm{a}} = n_{\mathrm{a}}) = \frac{(A_{\mathrm{t}}\lambda_{\mathrm{t}} + A_{\mathrm{o}}\lambda_{\mathrm{o}})^{n_{\mathrm{a}}}}{n_{\mathrm{a}}!} e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}} - A_{\mathrm{o}}\lambda_{\mathrm{o}}}$$

and subsequently assign every device to areas $\mathcal{A}_{\mathrm{t}}$ or $\mathcal{A}_{\mathrm{o}}$ through independent Bernoulli trials with probabilities

$$\Pr\left(\mathbf{U}_j \in \mathcal{A}_{\mathrm{t}}\right) = \frac{A_{\mathrm{t}}\lambda_{\mathrm{t}}}{(A_{\mathrm{t}}\lambda_{\mathrm{t}} + A_{\mathrm{o}}\lambda_{\mathrm{o}})} = \alpha \tag{16}$$

$$\Pr\left(\mathbf{U}_j \in \mathcal{A}_{\mathrm{o}}\right) = \frac{A_{\mathrm{o}}\lambda_{\mathrm{o}}}{(A_{\mathrm{t}}\lambda_{\mathrm{t}} + A_{\mathrm{o}}\lambda_{\mathrm{o}})} = 1 - \alpha, \tag{17}$$

respectively. This point of view relies on the splitting property of Poisson processes [25]. From Proposition 1, we gather that the maximum likelihood estimate for the rate of the aggregate Poisson process invariably fulfills $A_{\mathrm{t}}\lambda_{\mathrm{t}} + A_{\mathrm{o}}\lambda_{\mathrm{o}} = n_{\mathrm{a}}$. It then suffices to iteratively estimate α.

From the perspective of a single device, (16) and (17) represent the prior probabilities of being within or outside the target area. This same device can incorporate data vector $\mathbf{P}_j$ gathered by the sensing agents and compute its own posterior probabilities of being within or outside of the target area,

$$\begin{aligned}
\mathcal{Q}_{\mathcal{A}_{\mathrm{t}}}(j) &= \Pr\left(\mathbf{U}_j \in \mathcal{A}_{\mathrm{t}} | \mathbf{P}_j = \mathbf{p}_j\right) \\
&= \frac{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)\alpha}{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)\alpha + \mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)(1-\alpha)}
\end{aligned} \tag{18}$$

$$\begin{aligned}
\mathcal{Q}_{\mathcal{A}_{\mathrm{o}}}(j) &= \Pr\left(\mathbf{U}_j \in \mathcal{A}_{\mathrm{o}} | \mathbf{P}_j = \mathbf{p}_j\right) \\
&= \frac{\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)(1-\alpha)}{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)\alpha + \mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)(1-\alpha)}.
\end{aligned} \tag{19}$$

On the other hand, consider the reverse problem where every unit is aware of whether it lies within or outside the target area,

but model parameters are unknown. In this case, the system can compute the number of devices within the target area $r_{\rm t}$ and the number of devices outside the area $r_{\rm o}$. Using this information and the properties of the Poisson distribution, the maximum likelihood estimates for α, $\lambda_{\rm t}$, and $\lambda_{\rm o}$ become

$$\hat{\alpha} = \frac{r_{\rm t}}{r_{\rm t} + r_{\rm o}}, \qquad \hat{\lambda}_{\rm t}(r_{\rm t}) = \frac{r_{\rm t}}{A_{\rm t}}, \qquad \hat{\lambda}_{\rm o}(r_{\rm o}) = \frac{r_{\rm o}}{A_{\rm o}}.$$

Based on these observations, it is possible to build an iterative procedure akin to the expectation-maximization (EM) algorithm that yields estimates for the unknown model parameters and area occupancy $R_{\rm t}$. Initialization is derived from $A_{\rm t}$, $A_{\rm o}$, and the number of active wireless units,

$$\hat{\alpha}^{(0)} = \frac{A_{\rm t}}{A_{\rm t} + A_{\rm o}}, \quad \hat{\lambda}_{\rm t}^{(0)} = \frac{n_{\rm a} A_{\rm t}}{A_{\rm t} + A_{\rm o}}, \quad \hat{\lambda}_{\rm o}^{(0)} = \frac{n_{\rm a} A_{\rm o}}{A_{\rm t} + A_{\rm o}}.$$

After initialization, the algorithm seeks to find good estimates by iteratively alternating between the following two tasks.

1) Expected Locations: During this step of the iteration, we calculate the posterior probabilities that each wireless unit lies within or outside the target area. Under assumed model parameters $\hat{\lambda}_{\rm t}^{(k-1)}$ and $\hat{\lambda}_{\rm o}^{(k-1)}$, these probabilities are

$$\begin{aligned} \mathcal{Q}_{\mathcal{A}_{\rm t}}^{(k)}(j) &= \Pr\left(\mathbf{U}_j \in \mathcal{A}_{\rm t} | \mathbf{P}_j = \mathbf{p}_j; \hat{\lambda}_{\rm t}^{(k-1)}, \hat{\lambda}_{\rm o}^{(k-1)}\right) \\ &= \frac{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\hat{\alpha}^{(k-1)}}{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\hat{\alpha}^{(k-1)} + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\left(1 - \hat{\alpha}^{(k-1)}\right)} \end{aligned} \tag{20}$$

$$\begin{aligned} \mathcal{Q}_{\mathcal{A}_{\rm o}}^{(k)}(j) &= \Pr\left(\mathbf{U}_j \in \mathcal{A}_{\rm o} | \mathbf{P}_j = \mathbf{p}_j; \hat{\lambda}_{\rm t}^{(k-1)}, \hat{\lambda}_{\rm o}^{(k-1)}\right) \\ &= \frac{\mathcal{I}_{\mathcal{A}_{\rm o}}(j)\left(1 - \hat{\alpha}^{(k-1)}\right)}{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\hat{\alpha}^{(k-1)} + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\left(1 - \hat{\alpha}^{(k-1)}\right)}. \end{aligned} \tag{21}$$

We note that, at this point, one can sum the posterior probabilities and obtain the expected numbers of devices within and outside the target areas.

2) Likelihood Maximization: Given the expected number of active devices within and outside the target area, we can update the estimated value of model parameter α,

$$\hat{\alpha}^{(k)} = \frac{1}{n_{\rm a}} \sum_{j=1}^{n_{\rm a}} \mathcal{Q}_{\mathcal{A}_{\rm t}}^{(k)}(j) \tag{22}$$

This, in turn, leads to intensity parameters

$$\hat{\lambda}_{\rm t}^{(k)} = \frac{\hat{\alpha}^{(k)} n_{\rm a}}{A_{\rm t}} \qquad \hat{\lambda}_{\rm o}^{(k)} = \frac{\left(1 - \hat{\alpha}^{(k-1)}\right) n_{\rm a}}{A_{\rm o}}.$$

Although inspired by the expectation-maximization algorithm, this iterative procedure is not an instance of this famed algorithm [26]. Still, it is a valuable procedure to acquire estimates for the intensity parameters. Convergence is an important aspect of iterative schemes. For the method at hand, this property is established below.

Theorem 1: The iterative algorithm defined by alternating between Step III-C1 and Step III-C2 converges to a unique stable fixed point for any initialization condition $\alpha^{(0)} \in (0, 1)$.

Proof: See Appendix B. ∎

The fact that this iterative algorithm converges irrespective of $\alpha^{(0)} \in (0, 1)$ is a highly desirable attribute. Still, the performance of the algorithm depends on the ability of sensed signals to discriminate between prospective locations. In cases where the resolution of the observed power is low, performance may suffer. Having derived suitable estimators for our various inference problems, we turn to performance analysis. We first present simulation results based on our abstract model, and then continue with empirical data obtained through testbed implementation.

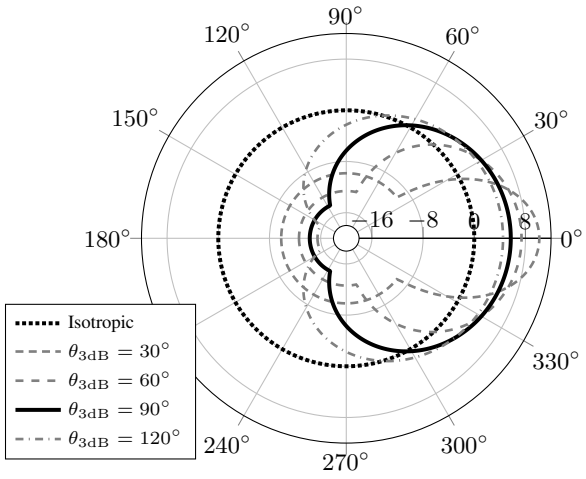

Fig. 2. This graph depicts normalized antenna radiation patterns for various 3 dB beamwidths. In all cases, the pointing direction is $\theta = 0°$ and the nominal attenuation floor is $G_{\rm floor} = 20$ dB.

IV. Numerical Simulations

In this section, we explore the potential benefits of directional antennas in the context of occupancy estimation. The target area $\mathcal{A}_{\rm t}$ is a square of dimension 6 m × 6 m inscribed in a larger square of dimension 10 m × 10 m. The two squares share a same center point, and the set difference between the two geometric shapes corresponds to the outside area $\mathcal{A}_{\rm o}$, as discussed in Section II.

Within our simulation framework, directional antennas are defined using an established 3GPP antenna model [27]. Each radiation pattern is characterized by a pointing direction of maximum gain and a beamwidth. Mathematically, we have

$$G_i(\phi_{ij}) = -\min\left\{12\left(\frac{\phi_{ij} - \theta_i}{\theta_{3\rm dB}}\right)^2, G_{\rm floor}\right\} - G_{\rm avg}$$

where $\theta_i \in (-180°, 180°]$, is the pointing direction (boresight) of the antenna attached to unit i; $\theta_{3\rm dB}$ is the 3 dB beamwidth of this radiation pattern in degrees; and $G_{\rm floor}$ is a nominal attenuation floor. The last variable $G_{\rm avg}$ is a normalization factor that enables a fair comparison between distinct antennas. The normalization constant, denoted $G_{\rm avg}$, ensures that all antennas radiate the same amount of power; its value is proportional to the average antenna gain

$$10\log_{10}\left(\int_{-180}^{180} \frac{10^{-\frac{1}{10}\min\left\{12\left(\frac{\phi_{ij}-\theta_i}{\theta_{3\rm dB}}\right)^2, G_{\rm floor}\right\}}}{360} d\phi_{ij}\right).$$

The antenna radiation patterns of candidate implementations are shown in Fig. 2.

The performance of an occupancy estimator is assessed through repeated trials. Findings are reported for isotropic antennas and for directional antennas with 3 dB beamwidth $\theta_{3\mathrm{dB}} = 90^\circ$, although general trends extend to other directional antennas. The sensing devices are located at the four corners of the target area, and they are pointing towards the common center of the squares. This configuration creates discriminating patterns for the directional antennas. For every active device located within the monitored area, a collection of four received signal strength measurements is acquired, one per sensing device. The aggregate information is subsequently employed in the inference task.

We generate every realization of the occupancy estimation problem as follows. Once λ_t and λ_o are set, the numbers of devices within and outside the target area are established using Poisson trials,

$$R_\mathrm{t} \sim \frac{(A_\mathrm{t}\lambda_\mathrm{t})^k}{k!} e^{-A_\mathrm{t}\lambda_\mathrm{t}} \qquad R_\mathrm{o} \sim \frac{(A_\mathrm{o}\lambda_\mathrm{o})^k}{k!} e^{-A_\mathrm{o}\lambda_\mathrm{o}}.$$

Each of the $R_\mathrm{t} = r_\mathrm{t}$ devices within the target area is assigned a location according to a uniform distribution, independently of other devices. Likewise, every device outside of the target area is placed at a location according to a uniform distribution. Altogether, this procedure yields $\underline{\mathbf{U}} = \underline{\mathbf{u}}$, the vector of realized locations for the wireless units. The final piece of randomness necessary to complete the trial is shadow fading. The components $\{L_{ij} = l_{ij}\}$ are generated independently following the log-normal distribution found in (2).

This information is subsequently transposed into the context of distributed sensing systems. The locations of the monitoring devices $\{\mathbf{s}_i\}$ along with their antenna radiation patterns $\{G_i(\cdot)\}$ are leveraged to compute the realized sequence of power vectors $\underline{\mathbf{p}} = (\mathbf{p}_1, \ldots, \mathbf{p}_{n_\mathrm{a}})$, where vector $\mathbf{p}_j$ and its entries are computed according to (1). The proposed estimators act on vector $\underline{\mathbf{p}}$ and they employ $\{\mathbf{s}_i\}$ and $\{G_i(\cdot)\}$ as side information.

To gain a better understanding of occupancy estimation, we examine two distinct scenarios and, within each scenario, we compare two alternate system configurations. The first setting is one where system parameters λ_t and λ_o are known. In this case, the Bayes estimator of Section III-A applies and performance can be assessed using (12), the Bayesian mean squared error. The second scenario is characterized by unknown system parameters. In this case, one can seek estimates for the system parameters λ_t and λ_o, and for the occupancy of the target region. This can be achieved by the approaches described in Section III-B and Section III-C. The average MSE is employed to characterized performance in the latter setting. Beside modeling assumptions, the system configurations can differ in antenna types. The base system features monitoring devices equipped with omnidirectional antennas, whereas the alternate configuration incorporates devices with directional antennas.

The physical parameters utilized throughout are derived from free-space path loss, regulation issued by the Federal Communications Commission (FCC), and profiles of typical wireless environments. Nominal power A is calculated using the Friis equation,

$$A = P_\mathrm{t} + 20\log_{10}\left(\frac{3\times 10^8}{f_\mathrm{carrier}}\right) - 20\log_{10}(4\pi)$$

where P_t is the transmitted power from the mobile devices, f_carrier is the frequency of operation for the Wi-Fi signals which is 2.462 GHz [28], [29], [30]. The logarithmic standard deviation, which characterizes variations in shadow fading, is set to $\sigma_\mathrm{s} = 2.0$ dBm. Pertinent values are summarized in Table I.

TABLE I
SYSTEM PARAMETERS USED DURING SIMULATIONS.

Physical Parameters	Values
Nominal Power	$A = -20.27$ dBm
Free-Space Loss parameter	$B = -20$ dBm
Logarithmic Standard Deviation	$\sigma_\mathrm{s} = 2.0$ dBm
3 dB Beamwidth (directional)	$\theta_{3\mathrm{dB}} = 90^\circ$
Antenna Floor	$G_\mathrm{floor} = 20$ dB

A. Bayes Estimation of Occupancy

We first consider the Bayes estimation framework introduced in Section III-A. We report our results using the Poisson splitting representation. That is, the aggregate Poisson rate across the two monitored regions is λ, and the splitting parameter between these two regions is α. This leads to known area-specific rates

$$\lambda_\mathrm{t} = \alpha\frac{\lambda}{A_\mathrm{t}} \qquad \lambda_\mathrm{o} = (1-\alpha)\frac{\lambda}{A_\mathrm{o}}.$$

Under this parameterization, we can plot performance results as a function of the splitting coefficient α. Each performance curve is associated with a specific value of $\lambda \in \{16, 32, 64\}$. The vertical axis represents the empirical BMSE of (12). A first set of curves in Fig. 3 showcases the BMSE of the occupancy estimator operating on data collected using isotropic antennas. The second group of curves, also shown in Fig. 3, corresponds to four directional antennas located at the four corners of the target area and pointing directly at the center. These antennas have a 3 dB beamwidth of $\theta_{3\mathrm{dB}} = 90^\circ$ and a nominal attenuation floor of $G_\mathrm{floor} = 20$ dB. Every point is obtained by averaging over fifty thousand trials.

The minimum mean squared estimator is algorithmically identical for cases with isotropic and directional antennas, and it takes into consideration antenna gains. Systems equipped with directional antennas perform uniformly better, irrespective of the aggregate rate λ and splitting coefficient α. This performance improvement is attributable to a more discriminating antenna configuration, which in turn leads to more informative observations.

B. Classical Estimation of Occupancy

In this section, we adopt the traditional viewpoint whereby the values of the parameters λ_t and λ_o are not known a priori. Within this context, it is meaningful to seek estimates for the Poisson rate parameters. Alternatively, one may wish to estimate the number of devices inside the target area. We

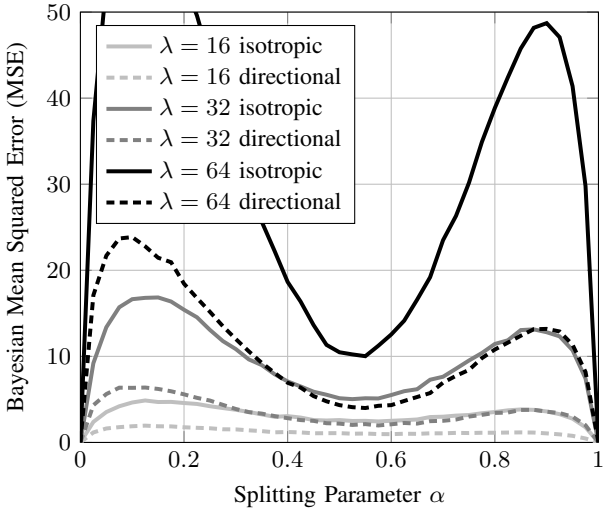

Fig. 3. This graph shows Bayesian mean squared error as functions of system parameters. Solid lines represent systems with isotropic antennas, whereas dashed lines correspond to the performance of systems with directional antennas.

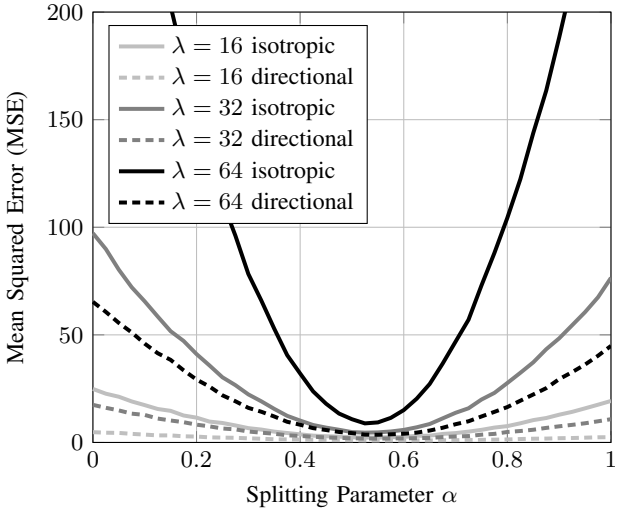

Fig. 4. This plot shows mean squared error in target occupancy as functions of splitting coefficient α under the maximum-likelihood scheme of Section III-B. The solid lines correspond to systems with isotropic antennas, whereas dashed lines are associated with systems using directional antennas.

examine the latter scenario, assessing performance using the average mean squared error. Section III proposes two solution paths to obtaining estimates: a maximum-likelihood estimator and a computationally efficient iterative algorithm. Figure 4 showcases the performance associated with the maximum-likelihood scheme for both isotropic and directional antennas. As before, the total Poisson rates studied are $\lambda \in \{16, 32, 64\}$ and the curves are functions of splitting coefficient α. We emphasize that the classical estimation problem at hand is inherently more difficult than the Bayes estimation task of Section IV-A, which explains the rise in mean squared error. This should be expected because there is more uncertainty in the present setting. More importantly, in the current scenario, a monitoring system entrusted with directional antennas too significantly outperforms the equivalent system with isotropic monitoring devices. Again, the more discriminating nature of directional antennas translates into sizable gains.

Similar characterizations are illustrated in Fig. 5 for the iterative procedure of Section III-C. Although numerically efficient, the iterative method is prone to error in noisy situations. When the gathered information is not very discriminating, e.g., devices with isotropic antennas, the iterative algorithm gives increasing weight to the larger of the two monitored areas, $\mathcal{A}_{\mathrm{t}}$ or $\mathcal{A}_{\mathrm{o}}$. In extreme cases where noise levels are very high, the iterative algorithm can converge to naive decisions where all the devices are estimated to be within the region of interest, or all of them are ascribed to the outside region. For these reasons, the iterative algorithm should only be employed in settings where the gathered information is sufficiently discriminating. Nevertheless, the iterative algorithm generally performs better with directional antennas. In some cases, the performance gains associated with narrow beamwidths are quite staggering. This provides further evidence to the benefits of using radiation patterns tailored to inference tasks in wireless monitoring systems.

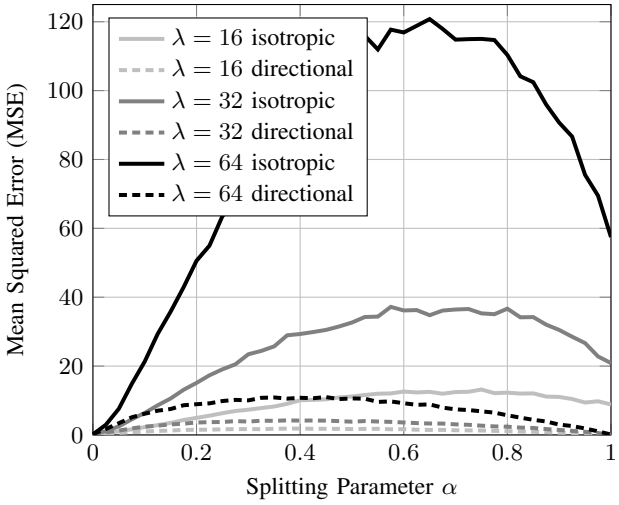

Fig. 5. This figure displays mean squared error in target occupancy as functions of system parameters for the iterative procedure of Section III-C. Solid lines represent systems with isotropic antennas, whereas dashed lines correspond to the performance of systems with directional antennas.

V. Testbed Implementation

To complement the numerical findings derived from our theoretical framework, we offer experimental results obtained through a testbed implementation. The prototype system is designed around Wi-Fi technology operating on the 2.4 GHz ISM radio band. In this experimental setting, wireless agents are connected to a designated wireless access point. The target area is a square of dimension 10 m $\times$ 10 m inscribed in a larger square of dimension 50 m $\times$ 50 m. A diagram showcasing the architecture for our monitoring system appears in Fig. 6.

1) Monitoring Devices: Every sensing device takes the form of a dedicated Next Unit of Computing by Intel™, and it runs an instance of GNU/Linux as its operating system. Wireless monitoring is enabled through an Alfa™ `AWUS036NHA`

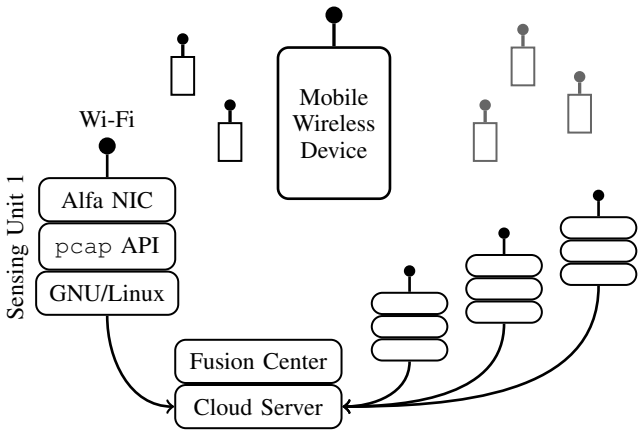

Fig. 6. Experimental results are based on four sensing units, each connected to a cloud server. The monitoring devices run the GNU/Linux operating system with Wi-Fi interfaces in monitoring mode. The inference tasks is to estimate the number of active agents (black) within an area of interest, and to distinguish them from agents (gray) outside of the target area.

wireless interface with a detachable antenna. This interface features an Atheros™ chipset which is capable of listening to all traffic present on a channel when operating in monitor mode. The reverse polarity subminiature version A (RP-SMA) antenna connector can be utilized to attach either an isotropic antenna or a directional custom patch antenna. A custom software build on the `pcap` application programming interface (API) captures, filters, and catalogs wireless packets. This software then creates a local database entry for every pertinent packet, which includes MAC address, received signal strength indicator (RSSI), and time. The information gathered by the monitoring devices is then relayed to a central server for post-processing.

2) Wireless Agents: To evaluate the performance of the proposed monitoring system, one needs to compare estimated results with a ground truth. In our testbed implementation, the ground truth is acquired by running a custom app on Android™ smartphones. Throughout the course of the experiment, every wireless unit runs this custom app, which logs GPS coordinates and time. The agents periodically transmit the information collected by the app to a central location. MAC addresses and time stamps are then used to match locations to power vectors at the inference center, yielding a data set for performance evaluation. Again, we emphasize that the proposed occupancy estimation algorithms can run without interaction with active units. The purpose of the app is simply to establish a ground truth about device location for benchmarking.

3) Experimental Samples: Altogether, the experimental set contains approximately 800 power and location vectors for monitoring devices with isotropic antennas, and another 800 power and location vectors for a system with directional antennas. Since there are four monitoring devices, this is equivalent to 6400 distinct points. Experimental trials are created by generating random subsets of the available data points. More specifically, a trial is created by first drawing Poisson random variables with parameters λ_{t} and λ_{o} for the numbers of devices within and outside the target area, respectively. Then, r_{t} entries are selected uniformly from agents in $\mathcal{A}_{\mathrm{t}}$; and r_{o} entries, from agents in $\mathcal{A}_{\mathrm{o}}$. The information corresponding to the resulting subsets is aggregated into a single vector $\underline{\mathbf{p}}$, which serves as input to the occupancy estimation algorithms. The estimates are then compared with the ground truth derived from the known locations.

The estimation schemes rely on knowledge of channel parameters and antenna gains. For the experimental section, antenna gains are measured empirically in an anechoic chamber. Channel parameters A and B can vary depending on the wireless environment [30], [31]. Our experimental profile can be categorized as suburban, and site-specific values for these parameters are obtained by applying the method of least squares to all the gathered data,

$$\underset{a,b}{\operatorname{argmin}} \left\| \underbrace{\begin{bmatrix} \vdots \\ p_{ij} - G_i(\phi_{ij}) \\ \vdots \end{bmatrix}}_{\mathbf{r}} - \underbrace{\begin{bmatrix} \vdots & \vdots \\ 1 & \log_{10}(d_{ij}) \\ \vdots & \vdots \end{bmatrix}}_{M} \begin{bmatrix} a \\ b \end{bmatrix} \right\|^2 .$$

Values for A and B are given in closed form by

$$\begin{bmatrix} A \\ B \end{bmatrix} = \left(M^{\mathrm{t}} M\right)^{-1} M^{\mathrm{t}} \mathbf{r}.$$

The variance σ_{s}^2 is computed using a standard unbiased sample variance estimator,

$$\sigma_{\mathrm{s}}^2 = \frac{1}{n-1} \sum_{i,j} \left(p_{ij} - A - B \log_{10}(d_{ij}) - G_i(\phi_{ij})\right)^2$$

where n denotes the sample size [22], [23]. The parameters for the isotropic systems are $A = -42.60$, $B = -21.97$, and $\sigma_{\mathrm{s}} = 8.64$ dBm. Similarly, the parameters for the systems with directional antennas are $A = -35.66$, $B = -24.83$, and $\sigma_{\mathrm{s}} = 10.15$ dBm.

Normality of the residual errors are tested using Shapiro-Wilk tests. Test values obtained are 0.9987 and 0.9979 for our isotropic and directional data sets, respectively. We note that, strictly speaking, errors cannot be Gaussian due to RSSI quantization. Nevertheless, a Shapiro-Wilk value close to one hints at the suitability of the Gaussian assumption, excluding quantization artifacts. This claim is further supported by looking at Q-Q plots comparing empirical quantiles to Gaussian quantiles, with all points lying close to the 45°-line as anticipated.

Experimental curves associated with occupancy estimation in the Bayesian setting appear in Fig. 7. The performance curves correspond to values of $\lambda \in \{16, 32\}$. As before, the horizontal axis denotes the Poisson splitting parameter α. The vertical axis indicates the experimental BMSE. Equivalent plots for the classical scenario where occupancy is assessed using a maximum-likelihood estimator for the intensity parameters, followed by the conditional expectation of (15), appear in Fig. 8.

In both cases, experimental findings exhibit trends similar to those identified through numerical simulations. Systems equipped with directional antennas are generally more discriminating and, hence, can performance substantially better. While

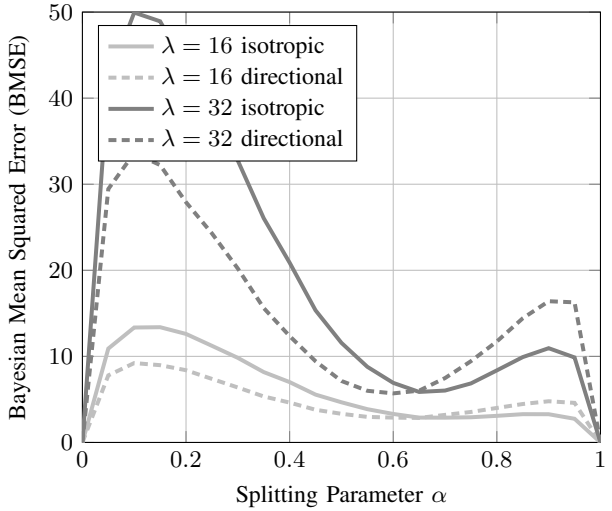

Fig. 7. This graph depicts the experimental Bayesian mean squared error as functions of Poisson splitting parameter α. Solid lines represent systems with isotropic antennas, whereas dashed lines correspond to the performance of systems with directional antennas.

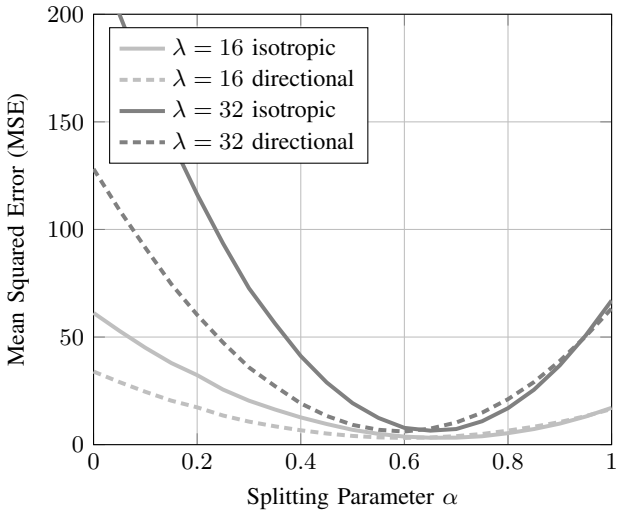

Fig. 8. This graph illustrates the experimental mean squared error in target occupancy as functions of splitting coefficient α for the classical scenario where Poisson intensity parameters are not known a priori. The solid lines showcase experimental findings with isotropic antennas, whereas dashed lines are for systems equipped with directional antennas.

comparing numerical simulations and experimental data, one should keep in mind a few considerations. The locations afforded by GPS units only form an approximate ground truth because readings contain inaccuracies. This can affect experimental results, especially for directional antennas where uncertainties in angles of incidence can produce large errors. This partly explains why the standard deviation σ_{s} is larger for the directional system when compared to its value for the implementation with isotropic antennas. Another source of modeling mismatch originates from the Android smartphones used during the data collection process. The theoretical results assume that wireless agents possess omnidirectional antennas whereas commercial smartphones have arbitrary radiation patterns [32]. This too can decrease performance, and it seems to affect systems with directional antennas more. Collectively, these issues point to the distinctions between theory and practice. Still, in most cases, experimental results corroborate the value of wireless antennas tailored to inference tasks. In systems engineered to take advantage of this design space, rather than systems constructed of existing components, one should expect gains to be even larger.

VI. Conclusion

This articles showcases how the antenna properties of a wireless sensing system can statistically improve the performance of inference algorithms. Our findings apply specifically to occupancy estimation and are based on Wi-Fi monitoring. However, this study hints at potential gains for other similar tasks, as radiation patterns are a common bridge between physical attributes and the gathered data. Such a framework is especially relevant for Wi-Fi devices because they periodically transmit probe signals to obtain information about nearby access points. These packets incorporate metadata about the identity of the senders, and they can be acquired passively using wireless sensing devices in monitor mode. Our research indicates that it is possible to accurately estimate the number of active agents within a prescribed area by deploying sensing devices about the area of interest, and that performance is generally enhanced by the careful shaping of antenna radiation patterns.

Conceptually, the radiation pattern of a sensing antenna acts as a spatial projection kernel that affects received signal strength. The performance of a monitoring system can then be enhanced by employing a configuration that strongly discriminates between wireless agents that are located within or outside the target area. This is illustrated in this article by positioning devices at the corners of the region of interest and comparing the performance of a system with directional antennas to that of the same system with omnidirectional antennas. In general, a more discriminating configuration yields significant improvements over a generic setup with isotropic antennas. Finding optimal configurations for specific scenarios is beyond the scope of this article. The optimal sensing configuration may depend on the geographic properties of the environment being monitored and the statistical profiles of the device population. However, one can expect performance to improve with the number of monitoring devices, but suffer in richer scattering environments.

Our findings point to several potential avenues for future research. This includes developing algorithms for tracking occupancy over time. It also encompasses the challenging task of assessing the benefits of reconfigurable and pattern dynamic antennas in the context of wireless inference. Agile antennas can be used to acquire very discriminating information about active devices, and warrants attention. Finally, several more complex statistical objectives can be cast within a similar framework. For instance, it should be possible to track the movement of a particular device within its environment, or infer interactions between devices based on proximity. Acquired information could be leveraged in adapting Wi-Fi

access point to current traffic conditions, or to implement preemptive caching for roaming devices.

APPENDIX A
PROOF OF PROPOSITION 1

In this section, we provide a proof for Proposition 1. This result asserts that the two-dimensional maximization of the likelihood function introduced in (13),

$$\max_{\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\geq 0} \mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)$$

reduces to a scalar optimization task,

$$\max_{\alpha\in[0,1]} \mathcal{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha,\frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha);\underline{\mathbf{p}}\right).$$

First, we note that the original likelihood function can be written as the product of an exponential function times a polynomial function in λ_{t} and λ_{o}. As such, this function is analytic and

$$\begin{aligned}&\max_{\lambda_{\mathrm{t}}\geq 0}\max_{\lambda_{\mathrm{o}}\geq 0} \mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)\\ &=\max_{\lambda_{\mathrm{o}}\geq 0}\max_{\lambda_{\mathrm{t}}\geq 0} \mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)\\ &=\max_{(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}})\in\Delta}\max_{c\geq 0} \mathcal{L}\left(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right),\end{aligned} \tag{23}$$

where Δ is the standard 1-simplex

$$\Delta=\left\{(t_1,t_2)\in\mathbb{R}^2\middle| t_0+t_1=1,t_0\geq 0,t_1\geq 0\right\}.$$

The next step in establishing Proposition 1 is to consider the maximization over scalar c. Specifically, suppose that parameters $(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}})\in\Delta$ are fixed. Then, under constant scaling, we get

$$\begin{aligned}&\mathcal{L}\left(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)\\ &=\sum_{k=0}^{n_{\mathrm{a}}}\frac{e^{-A_{\mathrm{t}}c\lambda_{\mathrm{t}}-A_{\mathrm{o}}c\lambda_{\mathrm{o}}}}{k!(n_{\mathrm{a}}-k)!}\\ &\times\sum_{\{\mathbb{I}\subset[n_{\mathrm{a}}]:|\mathbb{I}|=k\}}\left(\prod_{j\in\mathbb{I}}c\lambda_{\mathrm{t}}\int_{\mathcal{A}_{\mathrm{t}}}f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)d\mathbf{u}_j\right)\\ &\times\left(\prod_{j\in\mathbb{I}^{\mathrm{c}}}c\lambda_{\mathrm{o}}\int_{\mathcal{A}_{\mathrm{o}}}f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j)d\mathbf{u}_j\right)\\ &=c^{n_{\mathrm{a}}}e^{-(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})(c-1)}\mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right).\end{aligned}$$

Proceeding forward, we find the coefficient that maximizes this likelihood function. Since optima of unconstrained problems are found at critical points, we take the first derivative of the scaled likelihood function with respect to c,

$$\begin{aligned}&\frac{d}{dc}\left(\mathcal{L}(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}})\right)\\ &=\frac{d}{dc}\left(c^{n_{\mathrm{a}}}e^{-(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})(c-1)}\mathcal{L}(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}})\right)\\ &=n_{\mathrm{a}}c^{n_{\mathrm{a}}-1}e^{-(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})(c-1)}\mathcal{L}(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}})\\ &-c^{n_{\mathrm{a}}}e^{-(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})(c-1)}(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})\mathcal{L}(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}).\end{aligned}$$

Setting this derivative to zero, we get a unique solution

$$c=\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}}.$$

We perform a second derivative test to assess the character of this critical point,

$$\begin{aligned}&\frac{d^2}{dc^2}\left(\mathcal{L}(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}})\right)=\big(n_{\mathrm{a}}(n_{\mathrm{a}}-1)c^{n_{\mathrm{a}}-2}\\ &-2n_{\mathrm{a}}c^{n_{\mathrm{a}}-1}(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})+c^{n_{\mathrm{a}}}(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})^2\big)\\ &\times e^{-(A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}})(c-1)}\mathcal{L}(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}).\end{aligned}$$

Evaluating the second derivative at the critical point reveals that it is indeed a maximum,

$$\begin{aligned}&\frac{d^2}{dc^2}\left(\mathcal{L}(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}})\right)\Big|_{c=\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}}}\\ &=-\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}}\right)^{n_{\mathrm{a}}-2}n_{\mathrm{a}}e^{-n_{\mathrm{a}}+A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}}\mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)\\ &<0.\end{aligned}$$

Since the likelihood function vanishes when c approaches zero and infinity, we deduce that this maximum is unique. Applying this result to our original optimization problem in (23), we get

$$\begin{aligned}&\max_{\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\geq 0}\mathcal{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)=\max_{(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}})\in\Delta}\max_{c\geq 0}\mathcal{L}\left(c\lambda_{\mathrm{t}},c\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right)\\ &=\max_{(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}})\in\Delta}\mathcal{L}\left(\frac{n_{\mathrm{a}}\lambda_{\mathrm{t}}}{A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}},\frac{n_{\mathrm{a}}\lambda_{\mathrm{o}}}{A_{\mathrm{t}}\lambda_{\mathrm{t}}+A_{\mathrm{o}}\lambda_{\mathrm{o}}};\underline{\mathbf{p}}\right)\\ &=\max_{\alpha\in[0,1]}\mathcal{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha,\frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha);\underline{\mathbf{p}}\right).\end{aligned}$$

That is, the two-dimensional optimization problem reduces to a one-dimensional maximization.

APPENDIX B
PROOF OF THEOREM 1

This section offers a proof for Theorem 1. This is achieved through several steps. In Lemma 1, we show that the proposed algorithm has at most one fixed point in the open interval $(0,1)$. When this fixed point exists, we prove that its basic of attraction is the entire open interval in Lemma 2. A similar argument is employed in Corollary 1 to show that, when an interior fixed point does not exist, one of the end points act as an attractor. Altogether, these findings ensure that the iterative algorithm defined by alternating between Step III-C1 and Step III-C2 converges to the unique stable fixed point.

Lemma 1: Let $g(x)$ be the function obtained by combining (22) and (20), the two steps of the iterative scheme proposed in Section III-C,

$$g(x)=\frac{1}{n_{\mathrm{a}}}\sum_{j=1}^{n_{\mathrm{a}}}\frac{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)x}{\mathcal{I}_{\mathcal{A}_{\mathrm{t}}}(j)x+\mathcal{I}_{\mathcal{A}_{\mathrm{o}}}(j)(1-x)}. \tag{24}$$

The mapping $g:[0,1]\to[0,1]$ has at most one fixed point in the open interval $(0,1)$.

Proof: By definition, any fixed point of the iterative algorithm must fulfill the condition $x=g(x)$. Finding such fixed points is equivalent to identifying the zeros of

$$h(x)=x-g(x). \tag{25}$$

To study the properties of (25), we look at its first three derivatives

$$h'(x) = 1 - \frac{1}{n_{\rm a}} \sum_{j=1}^{n_{\rm a}} \frac{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\mathcal{I}_{\mathcal{A}_{\rm o}}(j)}{\left(\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)x + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)^2}$$

$$h''(x) = \frac{2}{n_{\rm a}} \sum_{j=1}^{n_{\rm a}} \frac{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\mathcal{I}_{\mathcal{A}_{\rm o}}(j)\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)}{\left(\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)x + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)^3}$$

$$h'''(x) = -\frac{6}{n_{\rm a}} \sum_{j=1}^{n_{\rm a}} \frac{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\mathcal{I}_{\mathcal{A}_{\rm o}}(j)\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)^2}{\left(\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)x + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)^4}.$$

For non-degenerate cases, the third derivative $h'''(\cdot)$ is negative for all points in $[0,1]$. As such, $h'(\cdot)$ is strictly concave over the same interval and it has at most two zeros within $[0,1]$. In turn, this implies that $h(\cdot)$ possesses at most two critical points. Since $h(0) = h(1) = 0$, we gather that $h(\cdot)$ has at most one zero in the open interval $(0,1)$. ∎

We proceed by examining the nature of the interior fixed point, if it exists.

Lemma 2: Suppose that a fixed point of $x = g(x)$ exists within the open interval $(0,1)$. Then, this fixed point is an attractor for the entire open interval.

Proof: We know from the proof of Lemma 1 that $h'(\cdot)$ is a strictly concave, continuous function. If $\alpha \in (0,1)$ is a fixed point of $g(\cdot)$, then $h(\alpha) = 0$. This, together with the fact that $h(0) = h(1) = 0$, implies that $h'(0)$ and $h'(1)$ are negative. Furthermore, since α must be located between the two critical points of $h'(\cdot)$, the derivative of $h(\cdot)$ evaluated at α must be positive and, consequently, $g'(\alpha) = 1 - h'(\alpha) < 1$. Moreover,

$$g'(\alpha) = \frac{1}{n_{\rm a}} \sum_{j=1}^{n_{\rm a}} \frac{\mathcal{I}_{\mathcal{A}_{\rm t}}(j)\mathcal{I}_{\mathcal{A}_{\rm o}}(j)}{\left(\left(\mathcal{I}_{\mathcal{A}_{\rm t}}(j) - \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)\alpha + \mathcal{I}_{\mathcal{A}_{\rm o}}(j)\right)^2}$$

is greater than zero. By continuity of $g'(\cdot)$, there exists a closed neighborhood of α, denoted $\mathcal{N}_\alpha$, such that $|g'(x)| < 1$ for all $x \in \mathcal{N}_\alpha$. Since $\mathcal{N}_\alpha$ is closed and bounded, it is compact and contains its limit-point. In particular,

$$\sup_{x \in \mathcal{N}_\alpha} |g'(x)| = B < 1.$$

By Taylor's theorem, $g(x) = g(\alpha) + g'(\xi)(x - \alpha)$, where ξ lies between α and x. Then, for any $x \in \mathcal{N}_\alpha$,

$$|g(x) - g(\alpha)| \leq B|x - \alpha|.$$

In other words, $g(\cdot)$ is a contractive mapping over $\mathcal{N}_\alpha$ and $\lim_{k\to\infty} g^k(x) = \alpha$ for any x in the closed neighborhood.

It remains to show that α is an attractor for the entire open interval $(0,1)$. To begin, assume $x \in (0,\alpha) - \mathcal{N}_\alpha$. Function $g(\cdot)$ is increasing therefore $g(x) \leq g(\alpha) = \alpha$. Also, $g'(\cdot)$ is strictly positive and $[0,1]$ is a compact set, therefore

$$M = \inf_{x \in [0,1]} g'(x) > 0.$$

Using Taylor's theorem one more time, we get

$$\alpha - g(x) = g(\alpha) - g(x) = g'(\xi)(\alpha - x) < M(\alpha - x).$$

In other words, every time the iterative algorithm is apply to $x \in (0,\alpha) - \mathcal{N}_\alpha$, it makes a step of minimal size towards α without exceeding it. The iterative sequence $g^k(x)$ is then bound to enter $\mathcal{N}_\alpha$ in a finite number of steps. A similar argument applies when $x \in (\alpha,1) - \mathcal{N}_\alpha$. Thus, starting with any point in $(0,1)$, the sequence $g^k(x)$ converges to the interior fixed point α. ∎

The same proof strategy can be applied to scenarios where there are no fixed points.

Corollary 1: Suppose that there does not exist an interior fixed point, then the iterative algorithm converges to zero or one.

Proof: This results is a simple extension of Lemma 2. Suppose there is no interior fixed point then either $h(x) > 0$ or $h(x) < 0$ for all $x \in (0,1)$. To begin, we adopt the former condition. First, we note that $g'''(\cdot) = -h'''(\cdot)$ and, as such, $g'(\cdot)$ strictly convex. If $g'(0) \geq 1$, then $g'(x) > 1$ for all $x \in (0,1)$ and

$$h(x) = x - g(x) = x - \int_0^x g'(\xi)d\xi \leq x - \int_0^x d\xi = 0,$$

which contradicts our original assumption. Hence, we have $g'(0) < 1$. Moreover, $g'(0)$ is greater than zero. By continuity of $g'(\cdot)$, there exists a closed neighborhood of zero such that $|g'(x)| < 1$ for all its members. Bounding the derivative over that interval and applying the contractive mapping theorem, we deduce that zero is an attractor over that neighborhood. For values of x within $(0,1)$ but outside the neighborhood, we use Taylor's theorem to show that the sequence $g^k(x)$ enters the neighborhood within finitely many steps. When $h(x) < 0$ for all $x \in (0,1)$, an analog argument applies, albeit with convergence to one. ∎

REFERENCES

[1] *Visual Networking Index: Global Mobile Data Traffic Forecast Update*, Cisco Systems, Inc., 2016.

[2] *IEEE Standard for Information technology – Telecommunications and information exchange between systems Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, LAN/MAN Standards Committee, IEEE Computer Society, 2012.

[3] N. Patwari and J. Wilson, "RF sensor networks for device-free localization: Measurements, models, and algorithms," *Proc. IEEE*, vol. 98, no. 11, pp. 1961–1973, 2010.

[4] J. Shen, A. F. Molisch, and J. Salmi, "Accurate passive location estimation using TOA measurements," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2182–2192, 2012.

[5] A. B. M. Musa and J. Eriksson, "Tracking unmodified smartphones using Wi-Fi monitors," in *Conference on Embedded Network Sensor Systems*. ACM, 2012, pp. 281–294.

[6] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin, "Measuring human queues using WiFi signals," in *International Conference on Mobile Computing & Networking*. ACM, 2013, pp. 235–238.

[7] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, "Signals from the crowd: Uncovering social relationships through smartphone probes," in *Internet Measurement Conference*. ACM, 2013, pp. 265–276.

[8] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 24–40, 2005.

[9] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, 2005.

[10] P. M. Djurić, M. Vemula, and M. F. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2229–2238, 2008.

[11] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[12] O. Ozdemir, R. Niu, and P. K. Varshney, "Channel aware target localization with quantized data in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1190–1202, 2009.

[13] J. Wang, R. K. Ghosh, and S. K. Das, "A survey on sensor localization," *Journal of Control Theory and Applications*, vol. 8, no. 1, pp. 2–11, 2010.

[14] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. K. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1495–1508, 2013.

[15] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, 2015.

[16] F. M. Naini, O. Dousse, P. Thiran, and M. Vetterli, "Opportunistic sampling for joint population size and density estimation," *IEEE Trans. Mobile Comput.*, vol. 14, no. 12, pp. 2530–2543, 2015.

[17] J. M. Smith and D. Towsley, "The use of queuing networks in the evaluation of egress from buildings," *Environment and Planning B: Planning and Design*, vol. 8, no. 2, pp. 125–139, 1981.

[18] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. ACM, 2010, pp. 1–6.

[19] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 17–29, 2002.

[20] X. Sheng and Y. H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 44–53, 2005.

[21] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Addison-Wesley, 1994.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993, vol. 1.

[23] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, 1998.

[24] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Duxbury Thomson Learning, 2001.

[25] S. M. Ross, *Stochastic Processes*, 2nd ed. Wiley, 1995.

[26] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[27] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, Technical Specification Group Radio Access Network, 3rd Generation Partnership Project, 2011, release 10.

[28] H. T. Friis, "A note on a simple transmission formula," *Proc. IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[29] J. D. Kraus, *Antennas*, 2nd ed. McGraw-Hill, 1988.

[30] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[31] A. F. Molisch, *Wireless Communications*, 2nd ed. Wiley, 2010.

[32] D. Tunon, J.-F. Chamberland, and G. H. Huff, "Orientation-awareness and wireless systems," in *Inf. Theory and App. Workshop (ITA)*. IEEE, 2015, pp. 230–234.