

Simultaneous Power-Based Localization of Transmitters for Crowdsourced Spectrum Monitoring

Mojgan Khaledi
University of Utah
mojgankh@cs.utah.edu

Sneha Kasera
University of Utah
kasera@cs.utah.edu

Mehrdad Khaledi
Rensselaer Polytechnic Institute
khalem@rpi.edu

Neal Patwari
University of Utah
npatwari@ece.utah.edu

Shamik Sarkar
University of Utah
shamik.sarkar@utah.edu

Kurt Derr
Idaho National Labs
kurt.derr@inl.gov

Samuel Ramirez
Idaho National Labs
samuel.ramirez@inl.gov

ABSTRACT

The current mechanisms for locating spectrum offenders are time consuming, human-intensive, and expensive. In this paper, we propose a novel approach to locate spectrum offenders using crowdsourcing. In such a participatory sensing system, privacy and bandwidth concerns preclude distributed mobile sensing devices from reporting raw signal samples to a central agency; instead, devices would be limited to measurements of received power. However, this limit enables a smart attacker to evade localization by simultaneously transmitting from multiple infected devices. Existing localization methods are insufficient or incapable of locating multiple sources when the powers from each source cannot be separated at the receivers. In this paper, we first propose a simple and efficient method that simultaneously locates multiple transmitters using the received power measurements from mobile devices. Second, we build sampling approaches to select mobile sensing devices required for localization. Next, we enhance our sampling to also take into account incentives for participation in crowdsourcing. We experimentally evaluate our localization framework under a variety of settings and find that we are able to localize multiple sources transmitting simultaneously with reasonably high accuracy in a timely manner.

1 INTRODUCTION

When software defined radios (SDRs) become ubiquitous, i.e., in the hands and pockets of average people, it will be easy for a selfish user to alter his radio(s) to transmit and receive data on unauthorized spectrum, for example, using an off-limits band, or transmitting/receiving on a channel when another device has priority. In addition, SDRs infected by a computer virus or malware could exhibit illegal spectrum use without the user's awareness. Cheap jamming

devices are already being used for illegal activities and this trend is likely to grow [3]. The U.S. Federal Communications Commission has an enforcement bureau which detects spectrum violations via complaints and extensive manual investigation. The mechanisms used currently for locating spectrum offenders are time consuming, human-intensive, and expensive. A violator's illegal spectrum use can be too temporary and mobile to be detected and located using existing processes. We envision a novel approach that crowdsources

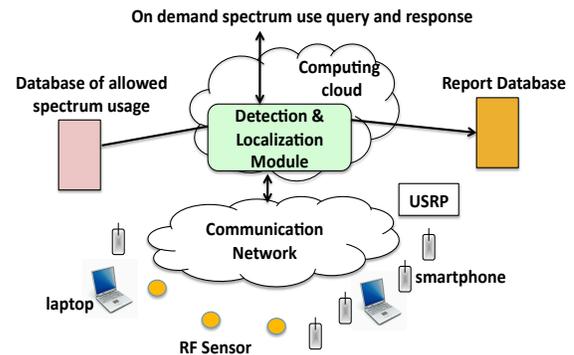


Figure 1: Enabling Distributed Set of Wireless Devices to Detect and Locate Spectrum Offenders.

the sensing and localization of spectrum offenders. We assume a distributed set of wireless devices, e.g., smartphones, radio frequency (RF) sensor nodes, laptops, access points and modems, etc., will participate by *sensing* the use of different bands of the spectrum over time and space and share their measurements with a detection and localization module on a (cloud) server, as shown in Figure 1. Our crowdsourced approach is expected to monitor a wide range of frequencies using a mix of SDR and non-SDR devices¹. The detection and localization module requests and collects spectrum usage information from a variety of sensors. It compares the spectrum usage with the allowed spectrum usage information (spectrum policies and regulations, frequency bands, locations, etc.) available in a database to determine and locate spectrum offenders. This

¹By 2020, the number of SDR devices is likely to be almost twice the number in 2014 [4].

spectrum usage database is akin to the whitespace database (e.g., an FCC-approved database that contains information on available whitespaces and their locations).

Towards the fulfillment of this vision, in this paper, we focus on crowdsourced localization of spectrum offenders. Due to privacy concerns and bandwidth and energy constraints, it is undesirable for mobile sensing devices to share raw signal samples with a central server, and hence, in our work, devices collect only received power measurements. Our mobile sensing devices do not decode any signals. A key challenge in localizing spectrum offenders is that of simultaneous localization of multiple transmitters². We develop a simple, yet efficient and accurate approach that simultaneously localizes multiple transmitters using the power of the sum of all signals received at the selected wireless devices. While localization of a transmitter based on power measurements taken by a network of sensors has been widely studied (see [30] for a survey), we argue that the existing work on WSNs cannot be simply adapted for localization of multiple transmitters using crowdsourcing. If two devices transmit at the same time, the received signal is a phasor sum of the signals from both. Simultaneous transmission in the same channel can be a consequence of an attacker’s violation of spectrum access rules or an intentional effort to jam. In either case, the signals may be impossible to separate, particularly when receivers report only power measurements to the server. Furthermore, it is important for crowdsourcing-based localization to account for the mobility and changing availability of user devices.

Many existing methods assume that if multiple transmitters are to be located, their signals can be separated at the receivers [30]. Note that even if all receivers were sophisticated enough to perform this blind source separation, a smart adversary could simply transmit signals that are not blind separable. When multiple signals cannot be separated, the few published methods [23, 25, 26] that are able to localize multiple transmitters from power measurements have high time complexity and do not consider the mobility and temporal availability of transmitters and receivers. For example, Quasi EM [26], a statistical approach to localize multiple transmitters, assumes that the transmitters and receivers are static and that the number of transmitters is known a priori. There is existing work on locating multiple transmitters using mobile robots [22]. However, this work is not applicable in our setting where mobile users move without network control. We need a localization algorithm which minimizes time complexity without significantly compromising the localization accuracy in dynamic environments in order to detect and locate unauthorized transmitters.

We present a method for simultaneous power-based localization of transmitters (SPLOT) for crowdsourced spectrum monitoring. We consider the temporal availability and mobility of both receivers and transmitters and make no assumptions about the number of transmitters. Our approach relies on the fact that, even when multiple transmissions overlap, typically the vast majority of power received by a receiver is from the nearest transmitter. Therefore, by finding local maxima in the spatially distributed received signal strength (RSS) measurements, we can approximate the region of presence of each transmitter. We can then convert the problem

of simultaneous multiple transmitter localization to a set of single transmitter localization problems and use a matrix inversion approach to find the location of each transmitter. Notably, we only consider an approximate region of each transmitter that is confined to the area around each local maximum and thus, achieve fast, accurate, and scalable localization.

Next, we build sampling approaches to select mobile sensing devices required for RSS measurements. Our goal is to select a set of wireless devices that provides maximum coverage for the monitored area considering mobility of both the sensing and the offending devices as well as possible erroneous or missed measurements. In this paper, we define and use a new metric called *degree expansion*, that represents the amount of overlap in the sensing ranges of mobile sensing devices. Using this metric, we propose two sampling approaches: 1) Greedy sampling, and 2) Metropolis sampling. While good Samaritans can be recruited for monitoring spectrum, mobile users need not participate in crowdsourcing for selfish reasons (including depletion of batteries and use of their processing resources) unless they receive some payoff as a compensation. We enhance our sampling approach to incentivize mobile users such that we select nodes that maximize coverage but minimize the total payoff. Furthermore, our incentive mechanisms motivate mobile sensing devices to act truthfully. Our *truthful sampling* considers both the budget limit and mobility of mobile sensing devices.

We experimentally evaluate our approach in two different settings: 1) an open environment with non-uniformly distributed receivers in the Orbit testbed [34] using USRP2 nodes for transmitting and receiving signals, and 2) a cluttered office with 44 uniformly distributed sensors [31]. Our experimental results show that using SPLOT we are able to localize multiple transmitters with high accuracy and in a timely manner. The highest average localization error using SPLOT measured in the open environment is 1.16 meters for up to 4 simultaneously transmitting transmitters, and the highest average localization error in the cluttered office with mobile transceivers is 2.14 meters. In comparison, the highest average localization error in Quasi EM measured in the open environment is above 6 meters. Our results also show that SPLOT is tens of minutes faster than Quasi EM. We also implement SPLOT on commodity devices and perform multi-transmitter localization in a variety of indoor and outdoor experiments. We find SPLOT to significantly outperform Quasi EM in these settings as well.

2 LOCALIZATION

In this section, we describe our approach to simultaneously locating multiple transmitters in the areas near the mobile sensing devices (also referred to here as receivers). The localization problem considered here is challenging for several reasons. First, there may be multiple transmitters, and the number of transmitters is unknown. Second, each measurement of received power is an unknown combination of the received powers from each transmitter. Third, the number and the locations of both transmitters and receivers might change from one sample to the next. Fourth, each link from transmitter to receiver experiences multipath fading, which is known to complicate RSS-based localization. We do not use time-of-arrival (TOA) methods because they require recording and sharing users’ sampled signals, which violates our privacy model. We do not use

²For instance, a malware-based attack could simultaneously cause many devices to violate rules or jam the spectrum.

angle-of-arrival (AOA) methods because they require additional RF hardware. We require an efficient and accurate localization approach that can locate all available transmitters in a dynamic environment using only power measurements.

2.1 Methodology

Our localization methodology assumes that receivers (mobile sensing devices) have been selected using our sampling approaches and that these receivers are spread geographically across the monitored area. Let K denote the unknown number of transmitters and $\theta = \{\theta_1, \theta_2, \dots, \theta_K\}$ represent their unknown two-dimensional locations. Our problem is to determine K and θ based on observed received powers reported by L receivers, $y = \{y_1, y_2, \dots, y_L\}$.

Our localization approach relies on two observations. First, receivers that are located near the transmitter observe generally higher power than the receivers that are distant from the transmitter. The second observation is that the observed RSS at each receiver is primarily affected by the nearest transmitter. To validate this observation, we compare the RSSs observed in selected receivers when there is no transmitter, when there is only one transmitter, and when we add multiple transmitters in different locations. Our results show that when there is a transmitter near a receiver, the RSS observed by this receiver increases substantially. However, there is a small growth on the observed RSS by the receiver when adding more transmitters at more distant locations.

These two observations allow us to reduce the problem of locating multiple, unknown number of transmitters to that of localizing a set of single transmitters as follows. First, we find the local maxima of RSSs observed by receivers that are greater than a predefined threshold. The predefined threshold is set to the minimum RSS that a receiver observes when there is a transmitter near it. By defining this threshold, we are able to separate the local maxima due to the presence of a transmitter from the local maxima due to the different fade levels at nearby receivers.

With the knowledge of the local maxima, our localization problem can be reduced to finding K transmitters, where K is equal to the number of local maxima. Instead of locating K transmitters in the entire monitored area, we divide the problem into K single transmitter localizations. For each local maximum, we locate a single transmitter. However, we confine the area to the small area around the local maximum and only use the RSSs that are observed by the receivers in this area for localization. Restricting the area to the small area around the local maximum reduces the time complexity of localization approach, and more importantly, confines the localization area and thus increases the accuracy of localization. This is because the noise in the measurements at receivers in other areas has little impact.

After reducing the multiple transmitter localization problem to a set of single transmitter localization, we locate each single transmitter in a small area around the local maximum using a single transmitter localization method. Specifically, we use a matrix inversion approach for single transmitter localization designed for higher efficiency and localization accuracy (see Section 5.3).

2.1.1 Single transmitter localization. In this section, we apply a linear model and perform inversion to localize the transmitted power. Our approach is designed to accurately locate a source with

unknown transmit power. In this linear model, we estimate the power field, i.e., the power transmitted vs. position. This power field is then used to find the location of an unknown transmitter.

Given a local maximum of RSSs observed by receivers, we find a transmitter located in a confined area around the local maximum. The confined area is a circle of radius R from the local maximum. Let vector $y = [y_1, y_2, \dots, y_L]$ be the received powers (in linear units of Watts) observed by L receivers located in the confined area. Also, let $x = [x_1, x_2, \dots, x_Q]$ be the power field, where x_i represents the transmit power sent by a transmitter in voxel i . We select a grid of Q voxels that fill the confined area. We have a forward model for y ,

$$y = Wx + n, \quad (1)$$

where n is an $L \times 1$ vector that represents the noise and fading contributing to the L RSS measurements, and W is an $L \times Q$ matrix where W_{ij} indicates the gain that would be experienced on the channel between a transmitter at voxel j and receiver i , if there is a transmitter in voxel j . The weight value W_{ij} is inversely related to the distance of the voxel and the receiver. We model the weight as,

$$W_{ij} = \begin{cases} d_{ij}^{-n_p}, & d > \text{minPL} \\ \text{minPL}^{-n_p}, & \text{otherwise} \end{cases} \quad (2)$$

Here, d_{ij} is the Euclidean distance from the center of voxel j to the receiver i . n_p is the path loss exponent and minPL is a minimum path length.

A power field estimate \hat{x} is estimated from y . However, finding \hat{x} is, in general, an ill-posed inverse problem. We use a regularized least squares approach to compute an estimate,

$$\hat{x} = \Pi y \quad (3)$$

$$\Pi = \left(W^T W + \sigma_N^2 C_x^{-1} \right)^{-1} W^T, \quad (4)$$

where σ_N^2 is the noise variance, W^T is the transpose of matrix W , and the prior covariance matrix C_x is obtained by using an exponential spatial decay function,

$$[C_x]_{jl} = \sigma_x^2 e^{-f_{jl}/\delta_c}. \quad (5)$$

Correlation distance constant δ_c describes the distance at which two voxels have correlation coefficient $1/e$, σ_x^2 is the variance of the transmit power field, and f_{jl} is the length of the line between the centers of voxels j and l .

Finally, the transmitter location is estimated to be the center of voxel with maximum value of \hat{x} .

2.1.2 Dynamic Localization. We also address the problem of locating mobile transmitters. We consider two dynamic cases: 1) the number and locations of transmitters are changing; and 2) the number and location of both transmitters and receivers are changing.

One approach to locate these dynamic cases is to repeat the multiple transmitter localization procedure every time new received power measurements are made. This approach is not efficient since changes are likely to be small from one time interval to the next. Thus, in this section, we present methods to use the coordinates from the previous estimate in the current localization problem.

Dynamic transmitters: When changes happen only in the number and location of *transmitters*, our dynamic localization works as follows.

The localization module uses three data sources at time t : 1) the RSS measurements of the selected receivers in time $t - 1$, which we denote y^{t-1} ; 2) the RSS measurements of the receivers in time t , y^t ; and 3) the locations of detected transmitters in time $t - 1$, which we denote θ^{t-1} . The first step is to estimate the number of transmitters, which we accomplish by comparing y^{t-1} with y^t . If there is no significant change, the localization module sets $\theta^t = \theta^{t-1}$. If the RSS increases significantly for some of the receivers, the localization module performs multiple transmitter localization for the area that is covered by these receivers to find new transmitters. If the RSS decreases significantly for some receivers, the localization module removes previous transmitters if there are receivers with significant decrease in the RSS within the confined area (circle of radius r) around the previous transmitters. The set of transmitters at time t , θ^t is the union of the remaining transmitters from the previous time and the newly added transmitters.

Note that by significant change, increase or decrease, in RSS, we mean any changes in the RSS that is greater than a predefined threshold. The predefined threshold is determined by the localization module depending on the environment. This predefined threshold is equal to the minimum change in the RSS measurements of receivers when at least one transmitter is added to or removed from the environment.

One drawback of the above approach is the propagation of error. Since the localization module estimates current transmitter locations using the previous results, the error can be carried forward from one time interval to the next, and after some time, the detected transmitters may be far from the actual transmitters. To deal with this problem, we periodically reinitialize by performing the multiple transmitter localization algorithm without any previous time estimates.

Dynamic transmitters and receivers: When the number and the location of receivers also change over time, we cannot compare two consecutive RSS measurements on the same link, as the change in receiver position will cause significant RSS changes. Even by mapping the receivers in the previous time, $t - 1$, to the nearest receiver available in time t and inversely relating the RSS to the distance between mapped receivers, we are not able to compare the RSSs in two consecutive time slots. This is due to the fact that RSS may be different for receivers with the same distance from the transmitter due to both shadowing and small-scale fading. Therefore, while the number or locations of receivers are changing, we perform multiple transmitter localization without using the previous time information. However, we limit the recalculation to at most every T seconds.

3 SAMPLING

An important aspect of offender localization is the task of selecting a set of mobile sensing devices for RSS measurements. The selection approach must consider the following: (i) maximum spatial coverage of the monitored area, and (ii) computational efficiency, i.e., select the nodes in a timely manner.

To select mobile sensing devices to maximize coverage, we define a new metric called *degree expansion*. A high degree expansion amounts to selecting receivers from new uncovered geographic areas and thus maximizing coverage.

Algorithm 1: Greedy Algorithm

```

input :  $G = (V, E, W)$ 
          $S$  // cardinality constraint
output:  $A$  // selected mobile sensing devices

1  $A = \emptyset$ ;
2  $v = \operatorname{argmax}_{v \in G} d_G(v)$ ; // the mobile sensing device with
   maximum degree is selected first
3  $A = A \cup v$ ;
4 while  $|A| < S$  do
5   Select new node  $v \in V - A$  where
6    $v \in \operatorname{argmax}_{i \in V - A_{j-1}} m_{i|A_{j-1}}$ 
7    $A = A \cup v$ ;
8 end

```

We develop two selection algorithms namely *Greedy* and *Metropolis*, to select a set of mobile sensing devices with high degree of expansion in a timely manner.

3.1 Preliminaries

We assume that the mobile sensing devices know their locations. Let T be the time interval (same as in Section 2) after which the sampling is repeated. At the beginning of each time interval T , the central controller sends the sensing request for that time interval to the mobile sensing devices. Then, mobile sensing devices voluntarily inform the central controller about their current locations and their sensing ranges. The sensing range of a mobile sensing device is the maximum radius around their current location for a given transmit power that they are able to measure. Given the current locations of mobile sensing devices, the central controller constructs a weighted graph $G = (V, E, W)$ where $v \in V$ represents a mobile sensing device (node), $e_{i,j} \in E$ denotes a link between node i and node j , and $w_{i,j} \in W$ represents the weight of link $e_{i,j}$. There is a link between node i and node j , $e_{i,j} \in E$, if their sensing ranges overlap. $w_{i,j} \in [0, 1]$ shows the amount of overlap between the sensing range of node i and node j and is obtained from the following formula:

$$w_{i,j} = \begin{cases} \frac{(r_i + r_j - d_{i,j})}{(r_i + r_j)} & d_{i,j} \leq r_i + r_j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Here, r_i and r_j are parameters related to the sensing ranges of node i and node j respectively. $d_{i,j}$ denotes the Euclidean distance between two nodes. We define the following:

- S is a set of nodes where $S \subset V$.
- $N(S)$ is the *neighborhood* of S that includes all nodes in $V - S$ that have a link to S .
- $d_G(v)$ is *degree* of node v in graph G . $d_G(v)$ is equal to sum of the weights of links that connect to v .
- The degree expansion, $DX(S)$, of a set S is $DX(S) = \sum_{i \in N(S)} \min_{j \in S} w_{i,j}$.

To select mobile sensing devices that maximize the coverage, we develop algorithms for solving the optimization problem below:

$$\arg \max_{S \subset V} DX(S) \quad (7)$$

Algorithm 2: Metropolis Algorithm

```
input :  $G = (V, E, W)$ 
         $S, M$  // cardinality constraint and number of
        iterations
output:  $A_{opt}$  // selected mobile sensing devices
// Initial sample,  $S$  nodes selected randomly.
1  $A_{current} = \text{random}(V, S)$ ;
2  $A_{opt} = A_{current}$ ;
3 for  $i = 1$  to  $M$  do
4    $v = \text{rand}(A_{current}, 1)$ ;
5    $w = \text{rand}(V - A_{current}, 1)$ ;
6    $A_{new} = \{A_{current} - v\} \cup \{w\}$ ;
7    $\alpha \in \text{rand}[0, 1]$ ;
8   if  $\alpha < \frac{q(A_{new})}{q(A_{current})}$  then
9      $A_{current} = A_{new}$ ;
10    if  $q(A_{current}) > q(A_{opt})$  then
11       $A_{opt} = A_{current}$ ;
12    end
13  end
14 end
```

where, S is the cardinality constraint. This optimization problem is NP-hard. It can be reduced to the *set cover* problem [14]. In the following sections, we propose two algorithms to obtain approximate solutions.

3.2 Greedy Algorithm

The greedy approach is shown in Algorithm 1. In the greedy algorithm, the selection is based on a greedy heuristic that selects the mobile sensing devices based on their marginal contributions to the degree expansion. The mobile sensing device with maximum degree expansion is selected first. Then, the mobile sensing devices are selected one by one iteratively based on their marginal contributions on degree expansion until S mobile sensing devices are selected. By iteration j , $j - 1$ mobile sensing devices are selected (denoted by A_{j-1}), and the marginal contribution of mobile sensing device $i \in V - A_{j-1}$, if represented by $m_{i|A_{j-1}}$ is equal to $d_G(i) - \sum_{j \in N(A_{j-1}) \cup A_{j-1}} w_{i,j} - \sum_{j \in A_{j-1}} w_{i,j}$. The mobile sensing device with maximum marginal contribution to the degree expansion is selected as the winner of the j iteration. I.e., $v \in \text{argmax}_{i \in V - A_{j-1}} m_{i|A_{j-1}}$. To simplify the notation, we replace $m_{i|A_{j-1}}$ by m_i in the rest of this paper.

3.3 Metropolis Algorithm

Metropolis is a Markov Chain Monte Carlo (MCMC) method to sample and evaluate probability distributions [24]. Recently, Metropolis sampling has been applied to subgraph sampling [18]. In this section, we use the idea of subgraph sampling to approximate the optimization problem of Equation 7.

Overview: Given a graph $G = (V, E, W)$, the idea of Metropolis algorithm is to create a subgraph of size $S < |V|$ in each iteration. The first set of mobile sensing devices is selected randomly and the subsequent sets of mobile sensing devices are constructed by removing a node from and adding a new node to the subgraph. We choose a *quality measure*, described below, to quantify the degree

expansion of samples. The acceptance or rejection of new mobile sensing devices is based on this quality measure. By selecting mobile sensing devices until convergence is achieved while keeping the selected mobile sensing devices with the maximum quality measure, we obtain a subgraph (selection of S mobile sensing devices out of N users) that approximately optimizes the degree expansion. We describe the pseudo-code of Metropolis in Algorithm 2.

Quality metric: Given a selected mobile sensing device A , the maximum possible degree expansion of graph $G = (V, E, W)$ is $\sum_{v \in V - A} d_G(v)$. This implies that the selected nodes, A , have links to all the other nodes, $V - A$ (i.e., $N(A) = V - A$). As a result, a normalized quality measure is equal to: $q(A) = \frac{DX(A)}{\sum_{v \in V - A} d_G(v)}$. The quality measure determines the degree expansion. A higher quality measure corresponds to a higher degree expansion.

4 TRUTHFUL SAMPLING

A naive way to provide incentives for the mobile sensing devices is to reward them such that the reward amount is greater than their declared cost of participation. The problem with this approach is that mobile sensing devices can over-report their costs and receive higher rewards. Therefore, the algorithm for selecting mobile sensing devices should also motivate them to truthfully report their costs. In this section, we enhance the greedy algorithm to consider both cost and degree expansion in selecting mobile sensing devices and add truthfulness. We design a time efficient payment mechanism for the case where the sampling is not optimal. Then, we propose the budget feasible version of the greedy algorithm. Finally, we propose a truthful mobility aware sampling approach that prevents selfish mobile sensing devices from cheating with regards to their availability for the sampling task. In truthful sampling, in addition to location and the measuring range, the mobile sensing devices also inform the central controller about their bids that can be equal or greater than their actual costs of collecting data, c_i .

Given graph $G = (V, E, W)$, and a cardinality or a budget constraint, the central controller selects winners and determines the payment, p_i , for each winner. We assume that all mobile sensing devices act rationally and selfishly, and their main goal is to maximize their own profits, not to harm others. Also, each user has a utility of $p_i - c_i > 0$ if selected and 0 otherwise.

4.1 Truthful Greedy Algorithm

Given the cardinality constraint S , the central controller wants to select S mobile sensing devices that maximize the degree expansion and minimize cost while considering individual rationality (provides the required incentives for selfish players to participate in the game) and incentive compatibility (truthfulness). This amounts to solving the following optimization problem.

$$\max_{S \in V} \sum_{i \in S} \frac{m_i}{c_i}$$

s.t.

$$\forall i, c'_i, p_i(c_i) - c_i \geq p_i(c'_i) - c_i \quad (8)$$

$$\forall i, p_i(c_i) - c_i \geq 0 \quad (9)$$

Here, Constraints 8 and 9 provide incentive compatibility and individual rationality.

Our allocation mechanism must consider both cost and the degree expansion. Our payment mechanism must ensure both incentive compatibility (IC) and individual rationality (IR). The best known payment mechanism for providing IC and IR is the well-known Vickrey-Clarke-Groves (VCG) mechanism [5]. Unfortunately, VCG fails to provide IC and IR when the allocation mechanism is not optimal. The problem of selecting a set of mobile sensing devices that maximize the total weight is equivalent to *set-cover* problem which is NP-hard. We use a Greedy algorithm to approximate the optimal solution. However, we still need a truthful payment mechanism that employs an approximated greedy algorithm to select mobile sensing devices. Towards this goal, we rely on Myerson's characterization of truthful mechanisms [21].

THEOREM 4.1. (Myerson 1981). *A mechanism with allocation mechanism A and payment mechanism P is truthful if and only if the following holds:*

- **A is monotone:** *The allocation mechanism keeps selecting the mobile sensing device i as a winner if it independently decreases its declared cost, c_i .*
- **P pays winners the threshold amounts:** *Paying each winner the maximum declared cost.*

First, we need to show the sub-modularity for $\frac{m_i}{c_i}$ which implies that:

$$\frac{m_1}{c_1} \geq \frac{m_2}{c_2} \geq \dots \geq \frac{m_S}{c_S} \quad (10)$$

We show the sub-modularity by contradiction. Suppose mobile sensing device i and mobile sensing device l are selected by the central controller where $i < l$, i.e., the mobile sensing device i is selected before the mobile sensing device l . Let us assume that $\frac{m_i}{c_i} < \frac{m_l}{c_l}$. According to the sampling mechanism, $[i] = \operatorname{argmax}_{j \in V - A_{i-1}} \frac{m_j}{c_j}$.

That means $\frac{m_{i|A_{i-1}}}{c_i} \geq \frac{m_{l|A_{i-1}}}{c_l}$. Given that $A_{i-1} \subset A_{l-1}$, $m_{i|A_{i-1}} \geq m_{l|A_{i-1}}$. Therefore, we have $\frac{m_{i|A_{i-1}}}{c_i} \geq \frac{m_{l|A_{i-1}}}{c_l} \geq \frac{m_{l|A_{l-1}}}{c_l}$, which is a contradiction with our assumption. As a result, Equation 10 is true. Next, we determine the payment using the Myerson's conditions. The key point is that we have to find the maximum value for the cost that a mobile node can declare and still win. We can find this threshold amount for each winner i by setting $V' = V - \{i\}$ and run the greedy algorithm until i is no longer selected. Let k be the index of the last mobile sensing device where $\frac{m_k}{c_k} < \frac{m_i}{c_i}$. Then, we obtain the payment to mobile sensing device i from the following formula:

$$p_i(c_i) = \max_{1 \leq j \leq k} \frac{c_j m_i^j}{m_j} \quad (11)$$

where m_i^j is the marginal contribution of mobile sensing device i on the degree expansion in iteration j .

LEMMA 4.2. *The payment mechanism provides individual rationality.*

PROOF. If a mobile sensing device is selected by the sampling mechanism, then the utility is equal to $\frac{c_j m_i^j}{m_j} - c_i$. Given that $\frac{m_i}{c_i} \geq \frac{m_j}{c_j}$, the utility of the winner is always positive. Also, if the mobile sensing device is not selected by the sampling mechanism, the utility of the device is zero. \square

LEMMA 4.3. *The payment mechanism provides incentive compatibility for the declared cost.*

PROOF. Let c_i, c'_i be the declared cost of mobile sensing device i when it is being truthful and not truthful, respectively. We must consider four cases here. First, if mobile sensing device i is the winner, it is still selected even by over-reporting, $c'_i > c_i$, or under-reporting, $c'_i < c_i$, its cost. In this case, the utility of the mobile sensing device decreases. Thus, the mobile sensing device has no incentive to lie. Second, if the mobile sensing device i is a loser it is still not selected even if it over-reports or under-reports its cost. Then, the utility is still zero and once again there is no incentive to lie. Third, if mobile sensing device i is a winner, is not selected by over-reporting its cost. Then, the utility will be zero. Thus, the mobile sensing device has no incentive to lie. Fourth, if the mobile sensing device i is a loser and it is not selected by over-reporting its cost. Otherwise, it contradicts the sub-modularity. However, the mobile sensing device can be selected by under-reporting its cost. Let assume that the mobile sensing device l is selected by the greedy algorithm when the mobile sensing device i acts truthfully. This means $\frac{m_l}{c_l} \geq \frac{m_i}{c_i}$. Since the maximum threshold for the mobile sensing device i is at least $\frac{m_l}{c_l}$, the payment to player i is at least $\frac{m_i^l c_l}{m_l}$ meaning that the utility is at most zero. \square

4.2 Budget-feasible Truthful Greedy Algorithm

We now consider a limit, B , on the total budget available to the controller. Given this limit, we must select a set of mobile sensing devices that maximize the marginal contribution on degree expansion. To select the mobile sensing devices with the budget limit constraint, similar to [38], we check the following condition at each iteration of the greedy algorithm, and stop the algorithm whenever the following condition no longer holds.

$$c_i \leq B \frac{m_i^j}{\sum_{j \in A_j} m_j} \quad (12)$$

Let $\{1, \dots, k\}$ be the largest subset that respects condition 12. To find the amount of payment for mobile sensing device $i \leq k$, we remove this player from the set of mobile sensing devices, $V' = V - \{i\}$, and then run the greedy algorithm. In each iteration of greedy algorithm, we check the budget limit constraint, Equation 12. Let k' be the last iteration of the greedy algorithm for $V' = V - \{i\}$ that Equation 12 holds and mobile sensing device i is still a winner, $\frac{m_{k'}}{c_{k'}} \leq \frac{m_i^{k'}}{c_i}$. To simplify the notation, we write the proportional

share of mobile sensing device i in iteration j , $\rho_i(j) = B \frac{m_i^j}{\sum_{j \in A_j} m_j}$ and $c_i(j) = \frac{c_j m_i^j}{m_j}$. The payment for user i in the budget feasible mechanism is:

$$p_i = \max_{j \leq k'} (\min(\rho_i(j), c_i(j))) \quad (13)$$

LEMMA 4.4. *The Budget feasible mechanism provides individual rationality.*

PROOF. Since the payment has the maximum value, we need to show the following for a $j \leq k'$.

- (1) $c_i \leq \rho_i(j)$

(2) $c_i \leq c_i(j)$

The first condition is the budget limit constraint, Equation 12, that holds. The second condition is related to the threshold payment and shows that the mobile sensing device i is still winner. Otherwise, the mobile sensing device i is not selected \square

LEMMA 4.5. *The budget feasible mechanism provides incentive compatibility for the declared cost.*

PROOF. To prove incentive compatibility, we need to show that p_i is the maximum cost that mobile sensing device i can declare and still be selected by the allocation mechanism. Let r be the index in $j \leq k'$ such that $\min(\rho_i(j), c_i(j))$ is maximized. For the case where $\rho_i(r) < c_i(r)$, declaring higher cost puts the mobile sensing device i after first k' users, therefore, the mobile sensing device i will not be selected by the sampling mechanism. Otherwise, if there is a j where $j \leq k'$ such that $\rho_i(j) > \rho_i(r)$, considering the maximality of r , $c_i(j) < \rho_i(r) < \rho_i(j)$. This implies that mobile sensing device i by increasing its cost gets placed after mobile sensing device j and thus, will not be selected. If $\rho_i(r) \geq c_i(r)$, a higher cost places the mobile sensing device i after r and since r is the maximum index in k' , mobile sensing device i will not be selected. Otherwise, if there is a j , $j < k'$ where $c_i(j) > c_i(r)$, by maximality we have $c_i(j) > \rho_i(j)$. i will not be selected because of the budget limit constraint, Equation 12. \square

4.3 Mobility Aware Budget-feasible Truthful Greedy Algorithm

In this section, we consider mobile sensing devices that can frequently change their locations. We propose a truthful mobility aware sampling that prevents the selfish mobile sensing devices from lying about their availability for the sensing task. In this setting, when the central controller probes the nearby mobile sensing devices for sampling over the time interval T , the mobile sensing devices reply by declaring their locations, costs, and their availability. Based on this information, the central controller selects a set of mobile sensing devices that maximize the marginal degree contribution under the budget limit and also are available during the time interval T . By providing incentives for the mobile sensing devices to declare the availability truthfully, the controller only selects among those devices that are available over the time interval T .

The sampling approach in the mobility-aware truthful mechanism is the same as one in the previous subsection. However, in the mobility aware approach, the central controller first selects the mobile sensing devices that are available during the time interval T and then applies the greedy budget limit on the available mobile sensing devices. The payment mechanism in this case should provide incentive for mobile sensing devices to declare their availability times truthfully. Therefore, in the mobility aware truthful mechanism, the threshold payment depends on cost and the availability time of mobile sensing devices. Let t be index of a time interval that mobile sensing device i participates in the sampling, then the payment to mobile sensing device i is the maximum of threshold payments for each time slot T . Formally,

$$p_i = \max_t(p_i(t)) \quad (14)$$

The above payment ensures the truthfulness of mobile arrival and departure times by removing the time dependency. Let a_i, d_i denote the arrival and departure times of device i , respectively. Let a'_i, d'_i be the declared arrival and departure time for the mobile sensing device i . Considering the fact that the mobile sensing device cannot declare early arrival and late departure, $[a'_i, d'_i] \in [a_i, d_i]$. Thus, the mobile sensing device i can only decrease the value of t . Since we take the maximum among all threshold payments, the payment to mobile sensing device i when it lies about the availability time is no greater than that when it reports truthfully.

4.4 Time complexity

Now, we show that truthful sampling can be done in polynomial time. For the Greedy algorithm, the time complexity is $O(S \log N)$. Recall that, S, N denote the cardinality constraints and the number of available mobile sensing devices, respectively. For the greedy truthful algorithm, the sampling time complexity is $O(S \log N)$. We also need to run the sampling mechanism S more times for payment determination. Thus, the time complexity is $O(S^2 \log N)$. This time complexity is the same for the budget-feasible algorithm. Finally, for the mobility-aware algorithm, the payment is the maximum threshold payment for each time slot. Therefore, the overall time complexity is $O(tS^2 \log N)$ where t is number of time intervals T that mobile sensing devices are available.

5 EXPERIMENTAL SETUP

To evaluate our sampling and localization approaches, we conduct experiments in two different areas: an open environment, and a cluttered office area. In this section, we describe these two areas and the mobility settings for transmitters and receivers. We also explain the evaluation metrics. The values of parameters used in our localization and sampling are listed in Table 1.

5.1 Test Environment

Open environment: In the open environment, there are no objects or obstructions in the area. This experiment is performed on the Orbit testbed [34], using the USRP2 nodes to transmit and receive signals. In this experiment, 14 receivers and 4 transmitters are placed non-uniformly inside of a 20 by 20 m area. The transmitters send a sinusoidal continuous wave (CW) signal, and the receivers measure received signal strength (RSS) at the same frequency.

Cluttered office: The office area is cluttered with desks, book-cases, filing cabinets, computers, and equipment. The collected data is a public data set [31]. In this experiment, 44 sensors are placed randomly in a 14 by 13 (m) area. Here, the transmitters are transmitting sequentially; and as this is a previously collected public experimental data set, we cannot change it to have multiple transmitters transmitting simultaneously. To approximate what the RSS would have been if multiple transmitters had transmitted simultaneously, we use the sum of linear received powers measured from each transmitter when transmitting sequentially. We justify this approximation as follows. First, we note that the expected value of the power of the sum of multiple signals is the sum of the powers of those individual signals [33]. Second, we perform a validation experiment in the Orbit testbed. For this validation, we randomly select two USRP2s and set them to transmit. We denote

Table 1: Evaluation parameters.

Parameters	Values	Descriptions
δ_p	1.5	Pixel width(m)
σ_x^2	0.5	Voxels variance(dB)
δ_c	1	Correlation coefficient
$minPL$	1.5	Minimum path length(m)
n_p	2	Path loss exponent
R	4	Confined area radius(m)
k	20	Number of iterations in Metropolis

the receiver linear power measurements as RSS(TX1) and RSS(TX2) when the transmitters transmit sequentially, and RSS(TX1,TX2) when they transmit simultaneously. Figure 2 shows the comparison of RSS(TX1)+RSS(TX2) vs. RSS(TX1,TX2). The data shows, in fact, that the linear approximation is very accurate, almost always within 0.5 dB.

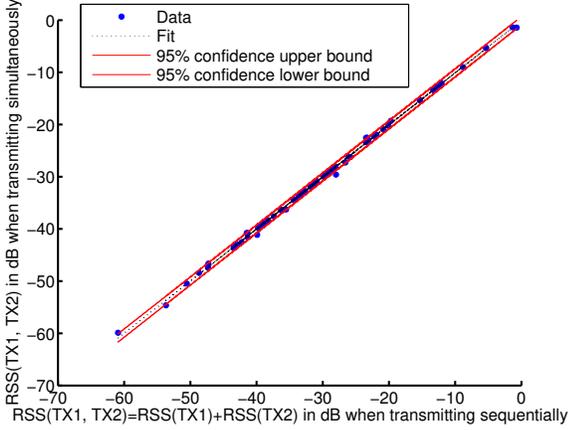


Figure 2: Correlation between the RSS received by receivers when transmitting simultaneously and when transmitting sequentially in the Orbit testbed.

Changes in Number and Locations of Transmitters and Receivers: To approximate mobility despite the fixed locations of mobile sensing devices in the two testbeds, we turn on and off the mobile sensing devices in the open environment and the cluttered office. To model the on and off states of each transmitter, we use a two state continuous time Markov chain that is able to model the bursts of device availability (on state) [29], with state 0 indicating off (or not transmitting) and state 1 indicating on (transmitting).

5.2 Evaluation Metrics

- (1) *Localization error:* The localization error, ϵ_l , is equal to the root mean squared error of the best assignment between estimated and actual transmitter locations [17].

$$\epsilon_l(\theta, \hat{\theta}) = \frac{1}{|\hat{\theta}|} \min_{a \in A} \left(\sum_{i=1}^{|\theta|} \sum_{j=1}^{|\hat{\theta}|} a_{i,j} d(\theta_i, \hat{\theta}_j)^2 \right)^{\frac{1}{2}} \quad (15)$$

where A is the set of all permutations between the set of actual transmitter locations θ and the set of estimated transmitter

locations $\hat{\theta}$, and $d(\theta_i, \hat{\theta}_j)$ is Euclidean distance between the i th actual and j th estimated transmitter locations.

- (2) *Cardinality error:* The cardinality error, ϵ_c , is the fraction of time during which the number of estimated transmitters differs from the actual number of transmitters [7].
- (3) *OSPA metric:* The optimal sub-pattern assignment (OSPA) metric, ϵ_p , penalizes the error in the number of estimated transmitters with a constant g that is measured in meters [36]. The OSPA metric is obtained by the following formula when $|\theta| \leq |\hat{\theta}|$,

$$\epsilon_p(\theta, \hat{\theta}) = \left(\frac{1}{|\hat{\theta}|} \min_{a \in A} \sum_{i=1}^{|\theta|} d^g(\theta_i, \hat{\theta}_{a_i})^2 + g^2 (|\hat{\theta}| - |\theta|) \right)^{\frac{1}{2}}, \quad (16)$$

where $d^g(\theta, \hat{\theta}) = \min(g, d(\theta, \hat{\theta}))$. When $|\theta| > |\hat{\theta}|$, the OSPA metric is obtained by inverting θ and $\hat{\theta}$ in (16).

5.3 Results

As mentioned in Section 2, after finding the local maxima, we perform localization in each sub area using the matrix inversion. This approach is simple and efficient in terms of time complexity. Based on our evaluation, the running time of the matrix inversion approach for a single transmitter in the open environment is 0.1 second, however, the running time of the well-known maximum likelihood estimator (MLE) [30] is around 1 second. In this section, we evaluate the matrix inversion approach in terms of the localization error. We compare the localization error of the matrix inversion approach with that obtained from the MLE. Table 2 shows the localization error in the open environment for each transmitter. As shown in this table, the matrix inversion approach preforms better in terms of the localization error. The average localization error of matrix inversion is 1.11 meters. In comparison, the average error of the MLE approach is 2.02 meters. Also, the variance of the localization error of the MLE approach is 4.1 square meters which is much higher than 0.18 square meters, the variance of the localization error of the matrix inversion approach. We also evaluate the localization error of one transmitter in the cluttered office. We find that the average and variance of localization error among all possible single transmitters are 1.6 meters and 0.61 square meters, respectively, in the matrix inversion approach. However, the average and variance of the localization error in the MLE approach are 1.77 meters and 1.55 square meters, respectively. Given the benefits of the matrix inversion approach in terms of both time complexity and the localization error, we select this approach to localize a single transmitter.

Next, we analyze the performance of SPLOT for multiple transmitters. To create the changes in number and locations of transmitters, we use the two state continuous time Markov chain and find the on and off states of each transmitter for 1000 seconds. The results are obtained by taking average of 100 different trials of transmitters in on and off states over 1000 seconds. We assume that only one transmitter is turned on at any instant, although once turned on at different time instants, multiple transmitters can be transmitting simultaneously. However, we do consider the scenarios where multiple transmitters can be simultaneously turned off.

Table 2: Localization error in Matrix inversion and MLE approaches for one transmitter in the open environment with no mobility.

Transmitter	Localization error(m)	
	Matrix inversion	MLE approach
TX1	1.41	1.01
TX2	0.5	1.03
TX3	1.11	0.97
TX4	1.41	5.08

Number of transmitters: We consider the impact of maximum number of transmitters on the performance of SPLOT. We set the maximum number of transmitters to 1, 2, 3, and 4 in the open environment. Figure 3 shows the changes in the average localization error of SPLOT and Quasi EM with the maximum number of transmitters in open environment. To find the localization error of Quasi EM, we provide the actual number of transmitters as an input parameter. Also to find the average localization error, we run 1000 trials of Quasi EM and take the average over these trials. The average localization errors shown in Figure 3 are over different combinations of transmitters in the open environment.

Figure 3 shows that the average localization error increases substantially with the increase in the number of transmitters in the Quasi EM approach. However, the average localization error of SPLOT is at or less than 1 meter, even though SPLOT also estimates the number of transmitters (unlike Quasi EM, SPLOT is not provided the number of transmitters as an input). Interestingly, we do not observe any penalty for increasing the number of transmitters being located.

Next, we compare the running time of SPLOT with that of Quasi EM in the open environment. Table 3 shows the running time of these two approaches for a trial of 100 seconds of the transmitters being in the on and off states. The Quasi EM is run only for one time. Table 3 shows that the running time of SPLOT is around 0.5 second. However, the running time of Quasi EM is above 200 seconds and the increase in the maximum number of transmitters substantially degrades the running time of Quasi EM degrades. We also expect the running time of Quasi EM to degrade more with increasing number of mobile sensing devices.

We also evaluate SPLOT in terms of average cardinality error and average OSPA error. Table 4 shows the average OSPA error and the average cardinality error of SPLOT with increasing maximum number of transmitters. This table shows that an increase in the maximum number of transmitters increases the average cardinality error. The average cardinality error in the worst case is around 0.14 when the maximum number of transmitters is 4. The OSPA metric considers both the localization error and the cardinality error in one metric. Table 4 shows that the average OSPA error increases by a very small amount even when the cardinality penalty is set to a very high value ($g = 5m$). This is because the fraction of times that the number of estimated transmitters $|\hat{\theta}|$ and the number of actual transmitters $|\theta|$ differ is very small.

Impact of transmitters' locations: Next, we analyse the effect of transmitter locations on the localization approach. To find the impact of transmitter locations, we use the cluttered office data

Table 3: Running time of SPLOT and Quasi EM in the open environment.

Maximum number of transmitters	Running time (second)	
	SPLOT	Quasi EM
1	0.5	211
2	0.5	985
3	0.7	2281
4	0.6	4169

Table 4: $\bar{\epsilon}_p(m)$, and $\bar{\epsilon}_c$, of SPLOT in the open environment.

Maximum number of transmitters	$\bar{\epsilon}_p(m)$			$\bar{\epsilon}_c$
	$g = 1$	$g = 2$	$g = 5$	
1	0.79	1.01	1.01	0
2	0.91	1.17	1.17	0
3	1.04	1.36	1.45	0.05
4	1.14	1.51	1.74	0.14

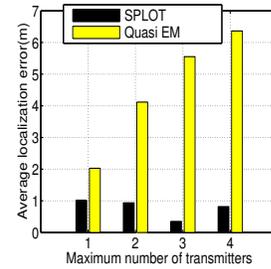


Figure 3: Average localization error versus the maximum number of transmitters in the open environment for Quasi EM and SPLOT.

with the maximum number of transmitters equal to 2. We consider different combinations of two transmitters such that the Euclidean distance between two transmitters varies from 3.5 meters to 18 meters. Figure 4 shows the relationship between the transmitters' Euclidean distance and the average localization error. Here, each data point is obtained by averaging over 100 different trials, each with its own randomly generated transmitter on and off chains. Also, note the different scales on the y-axes of the two plots. Figure 4 shows a linear correlation between the the transmitters distance and the localization error in the Quasi EM approach. However, in SPLOT, there is no apparent correlation between the distance between transmitters and the localization error. The correlation coefficients between distance and average localization error in the Quasi EM and SPLOT are 0.72 and -0.02 , respectively.

Regarding the cardinality error, there is a small correlation between the distance between transmitters and the cardinality error in SPLOT. Our evaluations show that the correlation coefficient of distance and average cardinality error in SPLOT is around -0.35 . With increasing distance, SPLOT is more successful in finding the local maxima and converting the multiple transmitters localization to a set of single transmitter localizations. Therefore, the average cardinality error decreases with increasing distance between transmitters. Similarly, there is also a small correlation of -0.1 between the distance and the OSPA error.

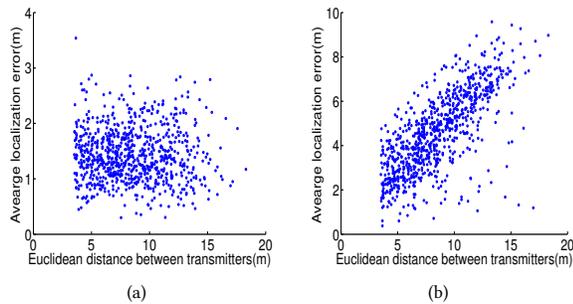


Figure 4: Impact of transmitters' locations in cluttered office (a) SPLIT, (b) Quasi EM.

Impact of number of mobile sensing devices: The number of available mobile sensing devices plays a key role in our localization approach. Selecting a large number of mobile sensing devices for measurement increases the communication overhead between the mobile sensing devices and the localization module. It also increases the time taken and the energy consumption. On the other hand, selecting a very small number of mobile sensing devices decreases the accuracy of our localization approach in terms of the localization error, the cardinality error, and the OSPA error. In this section, we examine the impact of the number of mobile sensing devices on efficiency of our localization approach and show how we can maintain the localization accuracy by using a suitable sampling approach to select mobile sensing devices among all available mobile sensing devices.

Table 5 shows the average localization error and the average cardinality error when the maximum number of transmitters is 3 in the open environment with the number of mobile sensing devices reduced from 14 to 12, 9, and 6 for three sampling approaches for selecting mobile sensing devices (Random, Greedy, and Metropolis). We make the following observations. First, the average localization error changes by about 0.25 meters when reducing the number of mobile sensing devices from 12 to 6 for all sampling approaches and the average localization error is close to 1 meter even for 6 mobile sensing devices (see Table 5). The Table 5 also shows that the average localization error does not change too much by selecting mobile sensing devices randomly. This table also shows that the average cardinality error increases with reduction in the number of mobile sensing devices. This is possibly because the number of selected mobile sensing devices is not enough to cover the whole area and the localization approach is not able to detect the transmitters located in the uncovered area. Moreover, the results of random sampling in Table 5 show that the location of mobile sensing devices directly effects the cardinality error. In other words, we need to have enough mobile sensing devices to cover the whole area to be able to detect all available transmitters. Third, the results of average cardinality error for Greedy sampling show that by reducing the number of mobile sensing devices and selecting a good sampling approach, we can still have a small cardinality error. Finally, this table also shows that for the open environment where nodes are distributed non-uniformly, the Greedy sampling preforms better than the Metropolis sampling approach.

Table 5: $\bar{\epsilon}_l(m)$, and $\bar{\epsilon}_c$, of SPLIT with different numbers of mobile sensing devices and three transmitters in the open environment.

Number of mobile sensing devices	Random		Greedy		Metropolis	
	$\bar{\epsilon}_l(m)$	$\bar{\epsilon}_c$	$\bar{\epsilon}_l(m)$	$\bar{\epsilon}_c$	$\bar{\epsilon}_l(m)$	$\bar{\epsilon}_c$
14	0.35	0.05	0.35	0.05	0.35	0.05
12	0.89	0.04	0.86	0.02	0.89	0.03
9	0.95	0.14	1.03	0.05	0.97	0.09
6	1.16	0.33	1.09	0.07	1.12	0.23

Table 6: $\bar{\epsilon}_l(m)$, and $\bar{\epsilon}_c$, with different numbers of mobile sensing devices and two transmitters in the cluttered office.

Number of mobile sensing devices	sensing devices	40	35	30	25	20	15	10	5
		Greedy	$\bar{\epsilon}_l(m)$	1.47	1.64	1.90	2.14	2.01	1.96
	$\bar{\epsilon}_c$	0.03	0.04	0.04	0.07	0.1	0.23	0.36	0.49
Metropolis	$\bar{\epsilon}_l(m)$	1.47	1.57	1.68	1.79	1.89	1.98	2.02	1.96
	$\bar{\epsilon}_c$	0.02	0.03	0.04	0.06	0.1	0.16	0.25	0.43

Next, we evaluate the impact of number of mobile sensing devices in the cluttered office. The maximum number of transmitters is 2 and we vary the number of mobile sensing devices between 5 to 40 with increments of 5. Table 6 shows average localization and cardinality errors in the cluttered office. We make the following observations. First, the average localization error varies from 1.42 to 2.24 meters. However, the average cardinality error increases relatively substantially with reduction in the number of mobile sensing devices. Second, for the cluttered office area where the mobile sensing devices are located uniformly in the environment, Metropolis sampling preforms slightly better than greedy sampling. Third, Table 6 shows that by reducing the number of mobile sensing devices from 42 to 25 and using either of Metropolis or greedy sampling approaches, we can keep the localization error close to 2 meters and the cardinality error to about 0.06. Fourth, the results of Table 6 shows that 5 mobile sensing devices are not enough to cover the whole area in the cluttered office environment. As this table shows, even by selecting the Metropolis sampling the average cardinality error is 0.43 if we only select 5 mobile sensing devices. As we discussed earlier, the determination of the number of mobile sensing devices required is done by the localization module. Generally, there is always a trade off between the accuracy of the localization algorithm, and the communication and processing overheads. Depending on the importance of each of these parameters and the localization results, the localization module makes a decision about the number of required mobile sensing devices.

Comparison of SPLIT and Quasi EM in the cluttered office: With the help of our experimental results, we have shown, earlier in this section, that SPLIT preforms better than Quasi EM in the open environment. However, the number of transmitters and mobile sensing devices in the open environment is small and there is no significant noise or obstruction in the environment. In this section, we compare SPLIT with the Quasi EM in the cluttered office environment, where we have 44 nodes that are located uniformly in the cluttered office and we can select any of these nodes as a transmitter or a mobile sensing device. Figure 5 shows the CDF of average localization error in the cluttered office for different

combinations of transmitters with maximum number of two transmitters. The average localization error on the x -axis corresponds to the localization error for each combination of transmitters obtained by averaging over 100 different trails. Figure 5 shows that the average localization error of SPLOT is significantly less than that of Quasi EM. Furthermore, the average localization error for any combinations of transmitters is 4.54 meters in Quasi EM which is much higher than the average localization error in SPLOT.

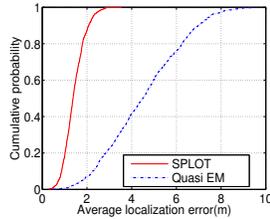


Figure 5: CDF of average localization error (m) in the cluttered office with maximum two transmitters.

6 IMPLEMENTATION

To further investigate its accuracy, we use commodity devices to implement SPLOT. Our mobile sensing device consists of a commodity smartphone/tablet that connects to an inexpensive Realtek dongle (RTL-SDR) [1] via a USB cable. The RTL-SDR acts as a mobile sensing device and collects raw In-phase/Quadrature(I/Q) samples. It operates in 25MHz-1750MHz with a sample rate up to 2.4MHz. We use BaoFeng BF-F8HP (BF) [2] transmitters that transmits VHF in 136MHz-174MHz and UHF in 400MHz-520MHz with up to 8W power. We build an Android smartphone/tablet app that measures spectrum in real time for a specified frequency range and sampling rate. Our app records the I/Q samples obtained from the RTL-SDR and computes the RSS values. In our set up, the app generates a (time, location, RSS) tuple every second. The location is the GPS coordinates for outdoor experiments. For indoor experiments, the app finds the location by indoor location fingerprinting [6].

Data Gathering. We have 30 users participate in our experiments for carrying both the transmitters and mobile sensing devices. Each user has their own Android mobile sensing device with our app installed on it. We collect data in different indoor and outdoor areas with at least two transmitters. The areas of our experiments are at least 30 m by 30 m³ in size.

To determine the location and the transmission time of transmitters, each user that carries a transmitter, also carries a mobile sensing device. Our app, on the device, records the transmission time and the transmitter location every second. In some experiments, the transmitters transmit continuously, while in the other experiments, we give the transmitters a transcript for transmission that shows the time of transmission for each transmitter. We create the transcript to allow us to experiment with different number of active transmitters at different times. In all experiments, we configure the transmit power to 1W and the frequency band to 443MHz.

³We make the experiment area small to evaluate SPLOT in an environment where sensing devices receive signals from both transmitters and the interference is strong.

Test Environment. We perform different indoor, outdoor experiments for both stationary and mobile scenarios.

Engineering Building- We perform two experiments on the third floor of an engineering building. In both experiments, there are at most 8 mobile sensing devices⁴ that are placed in four corridors of a square area of 40 m by 40 m. Two transmitters are located in two opposite corridors. In the first experiment (Experiment A), the mobile sensing devices are static while the transmitters transmit continuously and move along the corridors for 7 minutes at normal walking speed. In the second experiment (Experiment B), the mobile sensing devices move randomly in different corridors and the transmitters use a transcript for transmission.

Food Court- We perform two experiments in an indoor 30 m by 50 m university food court area. In both experiments, there are 6 mobile sensing devices located uniformly along the food court and both the transmitters and the mobile sensing devices are static. Also, in both experiments the transmitters use a transcript for transmission. In the first experiment (Experiment C), there are two transmitters that are located on two ends of the food court at first. Then, we gradually reduce the distance between the transmitters. In the second experiment (Experiment D), there are three transmitters located in three different corners of the food court.

Outdoor Area- We perform two experiments in an outdoor area of size 30 m by 50 m that is a part of a campus where both static (buildings, trees) and mobile obstacles (pedestrians) are present during the experiment. In both experiments there are at most 8 mobile sensing devices. In the first experiment (Experiment E), both transmitters and mobile sensing devices are static and the transmitters use a transcript for transmission. In the second experiment (Experiment F), both transmitters and mobile sensing devices are moving. The transmitters transmit continuously and move around a circle for 7 minutes at normal walking speed. The mobile sensing devices also move around the circle while maintaining a distance from the transmitters.

Results. We evaluate SPLOT using the data from our implementation⁵ and compare its performance with that of Quasi EM. Figure 6(a) shows the average localization error of SPLOT and Quasi EM in experiments A to F. Figure 6(a) shows that the average localization error of SPLOT is substantially less than the Quasi EM. The average localization error decreases in both SPLOT and Quasi EM when both transmitters and the mobile sensing devices are static (Experiment C, D, E). The average localization error of Quasi EM increases significantly when transmitters and mobile sensing devices are mobile (Experiment A, B, F). In comparison, the average localization error of SPLOT is less than 5 meters.

Table 7 shows the average OSPA error and the average cardinality error of SPLOT for experiments A to F. This table shows that the average cardinality error increases when both transmitters and mobile sensing devices are mobile. Also, the average cardinality error in Experiment D is greater than that in Experiments C and E because of an increase in the number of transmitters. Table 7 also shows that the average OSPA error increases by a very small amount even with high cardinality penalty ($g = 5m$).

⁴Users are not able to run the app for the duration of the experiment for different reasons such as a battery issue, etc.

⁵We are not able to evaluate our sampling approaches using this data because of the small number of mobile sensing devices.

Table 7: $\bar{\epsilon}_p(m)$, $g = 5m$, and $\bar{\epsilon}_c$, of SPLOT for experiments A to F.

Experiment	A	B	C	D	E	F
$\bar{\epsilon}_p(m)$	5.39	5.05	4.19	5.03	3.79	6.38
$\bar{\epsilon}_c$	0.18	0.11	0.04	0.16	0	0.06

Table 8: $\bar{\epsilon}_l(m)$, of SPLOT and Quasi EM versus distance between two transmitters in experiment C.

Distance (m)	45	41	36	29	18	9	6
SPLOT	3.38	3.82	3.67	4.92	2.9	3.68	1.16
Quasi EM	18.72	17.34	14.43	11.31	8.7	4.96	5.66

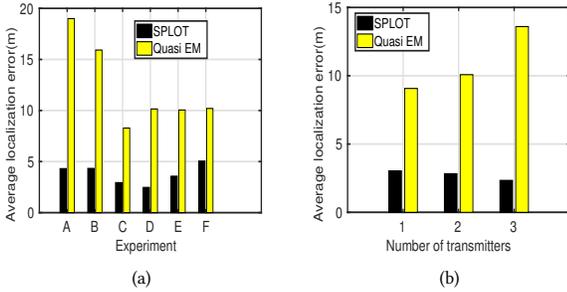


Figure 6: (a) Average localization error in experiments A to F (b) Impact of number of transmitters in the experiment D.

Figure 6(b) shows the changes in the average localization error of SPLOT and Quasi EM for different number of transmitters in Experiment D. Figure 6(b) shows that the average localization error increases with the increase in the number of transmitters in the Quasi EM approach. However, the average localization error of SPLOT is at or less than 4 meters, even though SPLOT also estimates the number of transmitters (recall that Quasi EM assumes that the number of transmitters is known).

Finally, we analyse the effect of transmitter locations on the localization approach using Experiment C with the number of transmitters equal to 2. We change the location of one transmitter such that the Euclidean distance between two transmitters varies from 6 meters to 45 meters. Table 8 shows the relationship between the transmitters' Euclidean distance and the localization error. Table 8 shows that there is no apparent correlation between the distance between transmitters and the localization error in SPLOT. However, the localization error decreases in the Quasi EM with decreasing distance between transmitters. This is because Quasi EM estimates the locations of both transmitters close to one of the transmitters.

7 RELATED WORK

Localization of multiple transmitters has been studied in existing work [22, 23, 25, 26, 30]. However, to the best of our knowledge, this paper is the first to develop and evaluate a localization approach that minimizes time complexity without significantly compromising localization accuracy in order to simultaneously locate unauthorized transmitters.

There are a great number of works that have focused on selecting a set of sensor nodes to provide the maximum coverage in wireless sensor networks (e.g., [8, 11, 39]). There are also a few existing works that consider both incentive and the coverage problem [15, 20, 40]. However, unlike our work, none of these existing works consider mobility, truthfulness, and the coverage problem all together. While [16] attempts to maximize the expected coverage for long-term participation, it does not apply to mobile environments where mobile sensing devices can be too temporary.

Recently, crowdsourcing using low-cost commodity radios has been used for spectrum monitoring (e.g., using RTL-SDR and smartphone [10, 28, 42], RTL-SDR and Raspberry Pi [32]), white space detection [35], and spectrum data decoding [9]. A few recent works on low-cost spectrum monitoring focus on transmitter identification [12, 43] and localization for detection of spectrum offenders [27, 41]. However, these existing works do not address the problems of multiple transmitters localization, sampling, and incentives, that we collectively address in our paper. The localization approach in [27] can only locate single static transmitters. However, our proposed localization approach, SPLOT, is able to locate multiple static or mobile transmitters. There is also some existing work on detection of spectrum misuse using only specialized hardware (e.g., USRPs [37] or spectrum analyzers [19]). However, this existing work, like [13], focuses on only monitoring spectrum usage and not on locating spectrum offenders.

8 CONCLUSION AND FUTURE WORK

We presented and evaluated a framework to locate multiple transmitters using crowdsourced measurements of received power. We addressed two main challenges in this framework. First, we presented a simple yet efficient and accurate method, SPLOT, for simultaneous localization of multiple transmitters using the received power measured by the selected mobile sensing devices. Second, we presented a sampling approach that determined the number and locations of required mobile sensing devices for measurement. Next, we enhanced our sampling to provide incentives for mobile sensing devices. We experimentally evaluated our framework and methods, and our results demonstrated the efficiency and accuracy of our approach.

Our work can proceed in the following directions. While evaluating our methodology, we have not explicitly evaluated how well the transmit power is estimated. We will perform this evaluation as future work. Very importantly, we will also investigate the effect of MIMO/beamforming as well as directional antennas on our localization methodology. Our approach assumes that mobile devices report RSS measurements and location information correctly. We will enhance our approach to make it robust against misreporting, malicious or erroneous, of RSS and device locations. While we use different smartphones and tablets, they all have the same RTL/SDR devices. In the future, we plan to study the impact of device heterogeneity on the quality of collected RSS measurements and its aggregation.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1564287.

REFERENCES

- [1] <http://sdr.osmocom.org/trac/wiki/rtl-sdr>.
- [2] <https://baofengtech.com/bf-f8hp>.
- [3] GPS Under Attack as Crooks, Rogue Workers Wage Electronic War. <http://www.nbcnews.com/news/us-news/gps-under-attack-crooks-rogue-workers-wage-electronic-war-n618761>.
- [4] Software-Defined-Radio Market Sees Increased Growth. <http://www.mwrf.com/software/software-defined-radio-market-sees-increased-growth>.
- [5] L. Anderegg and S. Eidenbenz. Ad hoc-VCG: A Truthful and Cost-Efficient Routing Protocol for Mobile Ad Hoc Networks with Selfish Agents. In *MobiCom*, 2003.
- [6] P. Bahl and V. N. Padmanabhan. RADAR: An In-building RF-based User Location and Tracking System. In *IEEE INFOCOM*, 2000.
- [7] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian. Multiple Target Tracking with RF Sensor Networks. *IEEE Trans. Mobile Computing*, PP(99), 2013.
- [8] J. Braams. A Node Scheduling Scheme for Energy Conservation in Large Wireless Sensor Networks. *Wirel. Comm. Mobile Comput*, 3(2), 2003.
- [9] R. Calvo-Palomino, D. Giustiniano, V. Lenders, and A. Fakhreddine. Crowdsourcing Spectrum Data Decoding. In *IEEE INFOCOM*, 2017.
- [10] A. Chakraborty, M. Rahman, H. Gupta, and S. R. Das. Specsense: Crowdsensing for Efficient Querying of Spectrum Occupancy. In *IEEE INFOCOM*, 2017.
- [11] B. Cărbunar, A. Grama, J. Vitek, and O. Cărbunar. Redundancy and Coverage Detection in Sensor Networks. *ACM Trans. Sen. Netw.*, 2(1), 2006.
- [12] A. Dutta and M. Chiang. "See Something, Say Something" Crowd-sourced Enforcement of Spectrum Policies. *IEEE Transactions on Wireless Communications*, 2016.
- [13] O. Fatemeh, R. Chandra, and C. Gunter. Secure Collaborative Sensing for Crowdsourcing Spectrum Data in White Space Networks. In *IEEE DySPAN*, 2010.
- [14] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 1998.
- [15] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos. Trac: Truthful Auction for Location-aware Collaborative Sensing in Mobile Crowdsourcing. In *IEEE INFOCOM*, 2014.
- [16] L. Gao, F. Hou, and J. Huang. Providing Long-Term Participatory Incentive in Participatory Sensing. In *INFOCOM*, 2015.
- [17] J. Hoffman and R. Mahler. Multitarget Miss Distance Via Optimal Assignment. *IEEE Trans. on Systems, Man and Cybernetics Part A: Systems and Humans*, 2004.
- [18] C. Hubler, H. P. Kriegel, K. Borgwardt, and Z. Ghahramani. Metropolis Algorithms for Representative Subgraph Sampling. In *ICDM'08*, 2008.
- [19] A. P. Iyer, K. Chintalapudi, V. Navda, R. Ramjee, V. N. Padmanabhan, and C. R. Murthy. SpecNet: Spectrum Sensing Sans Frontiers. In *NSDI*, 2011.
- [20] L. G. Jaimes, I. Vergara-Laurens, and M. A. Labrador. A Location-Based Incentive Mechanism for Participatory Sensing Systems with Budget Constraints. In *IEEE PerCom*, 2012.
- [21] C. Kim, D. Song, Y. Xu, J. Yi, and X. Wu. Optimal Auction Design. *Mathematics of operations research*, 1981.
- [22] C. Kim, D. Song, Y. Xu, J. Yi, and X. Wu. Cooperative Search of Multiple Unknown Transient Radio Sources Using Multiple Paired Mobile Robots. *IEEE Transactions on Robotics (T-RO)*, 2014.
- [23] R. Martin and R. Thomas. Algorithms and Bounds for Estimating Location, Directionality, and Environmental Parameters of Primary Spectrum Users. *IEEE Trans. Wireless Commun.*, 2009.
- [24] N. Metropolis, M. N. Rosenbluth, A. H. Teller, and E. Telle. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 1953.
- [25] A. Nasif and B. Mark. Collaborative opportunistic spectrum sharing in the presence of multiple transmitters. In *IEEE Globecom'08*, 2008.
- [26] J. K. Nelson, M. R. Gupta, J. E. Almodovar, and W. H. Mortensen. A Quasi EM Method for Estimating Multiple Transmitter Locations. *IEEE Signal Processing Letters*, 16(8), 2009.
- [27] A. Nika, Z. Li, Y. Zhu, Y. Zhu, B. Y. Zhao, X. Zhou, and H. Zheng. Empirical Validation of Commodity Spectrum Monitoring. In *ACM SenSys'16*, 2016.
- [28] A. Nika, Z. Zhang, X. Zhou, B. Y. Zhao, and H. Zheng. Towards Commoditized Real-time Spectrum Monitoring. In *HotWireless*, 2014.
- [29] J. Nonnenmacher, E. Biersack, and D. Towsley. Parity-based Loss Recovery for Reliable Multicast Transmission. *SIGCOMM Comput. Commun. Rev.*, 27(4), 1997.
- [30] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal. Locating the Nodes. *IEEE Signal Processing Magazine*, 22(4), 2005.
- [31] N. Patwari, A. O. H. III, M. Perkins, N. S. Correal, and R. J. OfiDea. Relative Location Estimation in Wireless Sensor Networks. *IEEE Trans. on Signal Processing*, 2003.
- [32] D. Pfammatter, D. Giustiniano, and V. Lenders. A Software-defined Sensor Architecture for Large-scale Wideband Spectrum Monitoring. In *IEEE/ACM IPSN'15*, 2015.
- [33] T. Rappaport. *Wireless Communications: Principles and Practice*.
- [34] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, and H. Kremo. Overview of the ORBIT Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols. In *IEEE WCNC*, 2005.
- [35] A. Saeed, A. Harras, E. Zegura, and M. Ammar. Local and Low-Cost White Space Detection. In *ICDCS*, 2017.
- [36] D. Schuhmacher, V. Ba-Tuong, and V. Ba-Ngu. A Consistent Metric for Performance Evaluation of Multi-Object Filters. *IEEE Trans. on Signal Processing*, 2008.
- [37] L. Shi, P. Bahl, and D. Katabi. Beyond sensing: Multi-GHz Realtime Spectrum Analytics. In *NSDI*, 2015.
- [38] Y. Singer. Budget Feasible Mechanisms. In *FOCS*, 2010.
- [39] G. Xing, X. Wang, Y. Zhang, C. Lu, R. Pless, and C. Gill. Integrated Coverage and Connectivity Configuration for Energy Conservation in Sensor Networks. *ACM Trans. Sen. Netw.*, 1(1), 2005.
- [40] X. Ying, S. Roy, and R. Poovendran. Incentivizing Crowdsourcing for Radio Environment Mapping with Statistical Interpolation. In *IEEE DySpan*, 2015.
- [41] T. Zhang and S. Banerjee. A Vehicle-based Measurement Framework for Enhancing Whitespace Spectrum Databases. In *ACM MobiCom*, 2014.
- [42] T. Zhang, A. Patro, N. Leng, and S. Banerjee. A Wireless Spectrum Analyzer in Your Pocket. In *HotMobile'15*, 2015.
- [43] M. Zheleva, R. Chandra, A. Chowdhery, A. Kapoor, and P. Garnett. Txminer: Identifying transmitters in Real-world Spectrum Measurements. In *IEEE DySPAN*, 2015.