2008

Accurate Localization of Low-level Radioactive Source Under Noise and Measurement Errors

Jren-Chit Chin

David K.Y. Yau
Purdue University, yau@cs.purdue.edu

Nageswara S. V. Rao

Yong Yang

Chris Y.T. Ma

Report Number:
08-012

Accurate Localization of Low-level Radioactive Source Under Noise and Measurement Errors

Jren-Chit Chin
David K. Y. Yau
Nageswara S. V. Rao
Yong Yang
Chris Y. T. Ma
Mallikarjun Shankar

CSD TR #08-012
May 2008

Accurate Localization of Low-level Radioactive Source Under Noise and Measurement Errors*

Jren-Chit Chin
Dept. of Computer Science
Purdue University
West Lafayette, IN
jcchin@cs.purdue.edu

David K. Y. Yau
Dept. of Computer Science
Purdue University
West Lafayette, IN
yau@cs.purdue.edu

Nageswara S. V. Rao
Oak Ridge National Laboratory
Oak Ridge, TN
raons@ornl.gov

Yong Yang
Dept. of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL
yang25@uiuc.edu

Chris Y. T. Ma
Dept. of Computer Science
Purdue University
West Lafayette, IN
ma18@cs.purdue.edu

Mallikarjun Shankar
Oak Ridge National Laboratory
Oak Ridge, TN
shankarm@ornl.gov

ABSTRACT

The localization of a radioactive source can be solved in closed-form using 4 ideal sensors and the Apollonius circle in a noise- and error-free environment. When measurement errors and noise such as background radiation are considered, a larger number of sensors is needed to produce accurate results, particularly for extremely low source intensities. In this paper, we present an efficient fusion algorithm that can exploit measurements from n sensors to improve the localization accuracy, and show how the accuracy scales with n. We report testbed results for a 0.911 μCi source to illustrate the effectiveness of our algorithm, in particular performance comparisons with state-of-the-art fusion algorithms based on Mean of Estimates (MoE) and Maximum Likelihood Estimation (MLE). We show that ITP is more accurate than MoE, whereas the choice between ITP and MLE is generally a tradeoff between accuracy and run time efficiency. Higher-intensity radioactive sources are not safe for actual experiments. In this case, we present simulation results based on a validated simulation model. We show that a low-intensity 400 μCi source, similar to the radioactivity of a concealed dirty bomb, can be localized to within 32.5 m using a sensor density of about 1 per 1100 m^2 in a surveillance area.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Source localization.

*Research supported in part by SensorNet program managed by UT-Battelle, LLC for U.S. Department of Energy under contract number DE-AC05-00OR22725, in part by the U.S. National Science Foundation under grant number CNS-0305496, and in part by a Purdue Research Foundation Fellowship awarded to J. C. Chin.

SenSys'08, November 5–7, 2008, Raleigh, North Carolina, USA.

General Terms

Algorithms, Design, Experimentation, Performance.

Keywords

Sensor network, radioactive source localization, sensor data fusion, noise and error management, iterative pruning.

1. INTRODUCTION

Urbanization is a powerful trend. In the U.S., more than 80% of the population will live in a city by 2025. The concentration of people increases their susceptibilty to stealthy attacks. When a radioactive stealthy bomb is detonated, for example, the small explosion might be dismissed as insignificant, but the real damage in terms of the number of people exposed to harmful radiation may be quite substantial. The hidden radiation can only be detected with suitable radiation sensors. At the same time, as sensor technologies mature which drive down costs and wireless connectivity becomes ubiquitous, it is possible to have dense in-situ networks of sensor nodes in urban areas for people protection. These sensors are embedded in the background and always on, and their measurements can be streamed wirelessly to a control center, where fusion of the data produces valuable information about important events.

In this paper, we are concerned with the near real-time localization of a low-level radioactive point source, similar in strength to a stealthy dirty bomb, in a geographical area. We assume that a network of n low-cost radiation sensors is available, where n can be large. Accurate and timely localization of the source would allow the source to be destroyed/removed or inform evacuation decisions for the people affected. Real-world deployments of such immersed radiation sensor networks are known. For example, the U.S. federal SensorNet project has deployed a network of RFTrax RAD-CZT [1] sensors in Washington D.C. to detect traces of gamma radiation in populated areas.

The localization problem has attracted a lot of interest in the sensor network community. When the direction to the source of the signal can be determined, e.g., acoustic or RF signals, three sensors are sufficient for localizing the source,

in an ideal deployment environment [12]. By ideal environment, we mean one in which noise and measurement errors can be filtered out or ignored in the localization process. In addition, a *range free* localization algorithm is proposed in [8] in which a node determines its position by measuring the RF signal strength from three anchor nodes and testing whether it is within the triangle formed by the anchors. The test is repeatedly performed for different groups of 3 anchors among all the $\binom{n}{3}$ anchors within communication range, and the node infers its position by finding a region where most triangles intersect. When determining the direction to the source of the signal is impossible, as in the case of gamma rays emitted by a radioactive source, three sensors will generally produce two estimates, each of which could be the real source position. In this case, a fourth sensor is sufficient to disambiguate the two estimates in an ideal environment. The solution uses a geometric approach based on the Apollonius circle [3].

Another widely used class of localization algorithms is based on the time difference of arrival (TDOA) [2, 18, 21]. TDOA (sometimes referred to as DTOA in the literature) measures the differences in time instants at which a signal emitted by the source reaches the different sensors. Hence, the exact distance between the source and a sensor is unknown, but the differences in distances between the source and each sensor can be inferred from the time differences. TDOA has been used for the localization of plumes (e.g., radioactive, biological, chemical) assuming idealized product-form plumes and an exponential decay function, i.e., a source of strength A will register a signal strength of Ae^{-r} at a distance r away [14]. When measurement and computational errors are considered, a geometry approach has been proposed [22, 17] to solve the TDOA problem with increased robustness. Their approach reduces the numerical instabilities when exact solutions of a system of quadratic equations are perturbed by noise. It is a general solution for TDOA problems, and is not specific to radiation localization.

In this paper, we show that given the specific focus of radiation localization, an inverse-square model is applicable and has been validated by our experiments, i.e., a source of strength A will register an intensity measurement of A/r^2 at a distance r away. In this case, the difference between two intensity measurements does not have a linear relationship with the difference in distances from each sensor to the source, and TDOA is not directly applicable. An adaptation of TDOA to log-space is suggested in [14, 15], in which TDOA is applied on the log of the intensity measurements, for use with radiation localization. Rather than adopt the log space transformation, our solution in this paper is to analyze the geometry based on taking direct ratios of square distances (RoSD) for each pair of sensors. We will present an experimental comparison between RoSD and log-space TDOA.

A novel approach to jointly solve the detection and localization problem for low-level radioactive sources is presented in [15]. In their approach, a tentative solution to the localization problem is solved by log-space TDOA using three sensors. After that, a sequential probability ratio test (SPRT) is applied to accept the solution, reject the solution as a phantom source due to background radiation, or return no decision due to insufficient information. The algorithm is designed to achieve given false alarm and missed detection ratios.

In real world deployments, noise and measurement errors are unavoidable. In this case, fusion of measurements by n sensors, $n > 3$, is useful to increase the robustness of the solution. For the distributed *detection* of radioactive sources, a copula method is proposed in [20] as the fusion technique. In their approach, each sensor performs a local hypothesis testing of whether the source is present or not. Based on the sensors' local decisions, a novel test based on copula theory and exploiting the correlated decisions of the sensors is then used to determine the final answer. For radiation source *localization* with n sensors, a mean of estimator (MoE) method is presented in [16] in which the localization results by each subset of three sensors are linearly combined to give the final answer. In [6], a maximum likelihood estimation (MLE) is proposed to search over a solution space of both the source strength and position, such that the difference between the predicted measurements according to the sensing model and the real measurements by the n sensors is minimized.

Our main contribution in this paper is to address the practical issues that arise in a near real-time radiation sensor network implementation using low-cost radiation sensors. In particular,

- We elucidate the sources of errors in an actual network, due to (1) background radiation and its inaccurate characterization; (2) variable efficiencies of different sensors and their incomplete determination; (3) the probabilistic nature of the underlying physical phenomenon causing statistical errors when the measurement time is limited, as required by the near real-time requirement; and (4) practical limitations of sensor designs and implementations.

- We show how the noise/errors can be mitigated using a sensor calibration process. In addition, we present a highly efficient iterative pruning (ITP) fusion algorithm for RoSD results produced by groups of three sensors among n available sensors. We show how the accuracy of the localization results improves with n, and compare the performance of ITP with those of the MoE and MLE algorithms. We show that ITP improves upon MoE by explicitly addressing the existence of *phantom estimates* computed by RoSD and by preferring higher-SNR non-phantom estimates in the fusion process. We show that ITP is somewhat less accurate than MLE, but generally runs much faster because it does not require an expensive optimization step over a large solution space.

- For an extremely low level 0.911 μCi radioactive source, we present actual testbed results for the performance evaluation. In addition, we present simulation results using a validated simulation model so that we can systematically explore the parameter space (e.g., vary the source strength) that is impossible with the actual testbed. For a higher- (but still low-) intensity source, comparable to the intensity of a stealthy dirty bomb, experimentation is not safe. In this case, we present a larger scale simulation based on the validated simulation model, and show that such a source can be localized in near real-time and with good accuracy in practice.

The balance of the paper is organized as follows. In Section 2, we formulate our problem, with explicit account for

the background radiation and sensor sensitivities in the sensing model. The ratio-of-square-distance (RoSD) localization algorithm using measurements by three sensors is presented in Section 3. We also define the notion of *phantom estimate* when RoSD produces two equivocal solutions. When noise and measurement errors are considered, fusion of measurements by n sensors is necessary. In Section 4, we review two state-of-the-art fusion algorithms based on mean of estimates (MoE) and maximum likelihood estimator (MLE). We then design an *iterative pruning* algorithm that (1) aims to explicitly eliminate phantom estimates in the fusion process, and (2) prefers higher SNR non-phantom estimates in producing the final result. Testbed results and simulation results using a validated simulation model are presented in Section 5 to illustrate the performance of the algorithms. In particular, we evaluate the impact of different sources of errors on the localization accuracy. We also compare ITP with MoE and MLE in terms of accuracy and efficiency. Simulation results are presented in Section 6 for higher intensity radiation sources similar to concealed dirty bombs. Section 7 concludes.

2. PROBLEM FORMULATION

We consider the localization of a static point radioactive source of unknown strength A_0 in a two dimensional surveillance area $\mathbb{R}^2$. Such a source might correspond to a dirty bomb left at a concealed location in a train station or other public venues. The unknown location of the source is given by $(x_s, y_s) \in \mathbb{R}^2$. The source induces a radioactivity of intensity $A(x, y)$ at any location $(x, y) \in \mathbb{R}^2$. Let S_i denote a radiation sensor located at $(x_i, y_i) \in \mathbb{R}^2$. Background radiation is universally present, and its measurement by S_i is denoted as B_i. All quantities of radioactivity are in terms of the number of emitted radioactive particles per time interval and are given in counts per minute (CPM). The objectives of the localization algorithm are to compute (1) an estimate of the radioactive strength $\hat{A}_0$, and (2) an estimate of the source location $(\hat{x}_s, \hat{y}_s)$, such that the difference between the estimate and the corresponding true value is minimized.

2.1 Radiation sensing model

Existing sensing models for radiation (e.g., [15]) consider background radiation as a point source whose radioactivity is superimposed on that of the unknown point source to be localized. They typically assume ideal sensors, and do not account for the effects of *sensor efficiency* on the sensor readings. Real radiation sensors, however, have different sensitivities, and their CPM readings may differ significantly even when they are subjected to the same level of radioactivity. Converting the CPM readings by sensors to the true underlying radioactivity levels, thus accounting for the *measurement errors* in the sensing process, requires therefore a careful calibration process which has not been considered in [15].

We propose a radiation sensing model to account for the efficiency of real sensors and hence their measurement errors. We note that a typical radiation sensor uses a detector to record the amount of radioactivity occurring over a given time interval, say one minute, and reports the recorded values per time interval. In the United States, the popular radiation measure is the number of disintegrations of the radioactive material in question over a time period. The measurement unit is Curie (Ci), where 1 Ci corresponds to 37 billion disintegrations per second, and 1 μCi corresponds to 2.22 million disintegrations per minute (DPM). Depending on the technology and the exposure surface of its detector, a radiation sensor will only detect a fraction of the disintegrations. We denote the fraction of the disintegrations detected by sensor S_i as the sensor efficiency E_i, of unit CPM/DPM. We assume that the radiation intensity decays over space according to the inverse square law [15]. Under this model, the radiation intensity due to a point radioactive source is given by $A(x, y) = \frac{A_0}{(x_s-x)^2+(y_s-y)^2}$. Consequently, the CPM reading measured by sensor S_i is given by

$$I_i = \frac{2.22 \times 10^6 \times A_0 E_i}{(x_s - x_i)^2 + (y_s - y_i)^2} + B_i. \quad (1)$$

The sensor efficiency E_i is assumed to be a constant. It is determined empirically with a calibration procedure. During the calibration, we need to first determine the background radiation intensity. To do so, the radiation sensor is placed away from any known radioactive source, and its intensity readings are collected over a long period of time. The average of the collected readings is then used as the estimate of the (average) background radiation intensity B_i. After B_i is determined, a low-level radioactive source of known strength A_0 is placed directly on top of the radiation sensor to be calibrated, and the intensity readings reported by the sensor are again collected over a long period of time. The average of these readings is then used as the measurement I_i for the source intensity A_0. Using these values, the sensor efficiency is computed by $E_i = \frac{I_i - B_i}{2.22 \times 10^6 \times A_0}$. Note that even sensors of the same model from the same manufacturer may have different E_i values. Hence, calibration should be ideally performed for each individual sensor.

Lastly, we note that the above model applies to all three types of Alpha, Beta, and Gamma radiation. Our application context in this paper is, however, Gamma radiation because it is highly penetrating and therefore the most dangerous type of radiation among the three.

2.2 Sources of noise and errors

The localization of a radioactive source, especially one of low intensity, is hard due to the nature of radioactive disintegrations, the presence of background radiation, and limitations in existing sensing hardware. *First*, the Gamma rays emitted by a point source (such as CS-137) are probabilistic and follow the Poisson process [10, 11, 15]. Because of the randomness of the phenomenon being observed, a single instantaneous reading of the source intensity is not reliable in determining the true source strength. This is another form of measurement error, which we call the *statistical error*, and affects the accuracy of A_0 used in our radiation model. In principle, a perfect measurement would require an infinite measurement time interval. In practice, only a limited number of samples can be used, particularly in the case of realtime or near-realtime localization. The magnitude of the statistical error is given by the standard deviation of the difference between the measured intensity and the true intensity, and is proportional to $\frac{1}{\sqrt{m}}$ where m denotes the number of samples. In localizing a low-level radioactive source, whose intensity measurement is heavily influenced by the background radiation, a sensing time on the order of minutes is required for COTS radiation sensors such as the RFTrax. The sensing time can be shortened with higher efficiency sensors.

Second, background radiation is universal, but is extremely hard to characterize due to many variables of natural and man-made radioactive sources. Examples of natural radioactive sources include potassium (K-40) radio-isotopes present in natural foods like bananas, carbon (C-14) in vegetation and other organic materials, and cosmic rays in space. Examples of man-made radioactive sources include smoke detectors, X-ray machines, and nuclear fuels. We are exposed to any combination of these common sources in everyday life, which is not a cause for concern and hence should be classified as part of the background radiation B_i. However, because each source is probabilistic and their exact combinations are not known and vary with time and space, a perfect characterization of the background radiation is impossible. While the calibration procedure described above can mitigate the effects of the background radiation to a significant extent, we have to acknowledge the presence of the background radiation as an unavoidable source of *noise* that will affect the localization task.

Third, limitations of sensing technologies and cost considerations for the sensing hardware may affect the accuracy of measurements. This is explained above as the sensor efficiency E_i, and is a source of *measurement errors* in the sensing process. While calibrating sensors to determine their E_i values can mitigate the effects of measurement errors to a significant extent, dust and other contaminants gathering on the surfaces of detectors over time may degrade the sensor efficiency. In this paper, we acknowledge that the calibration process is not perfect, and sensors may not always be re-calibrated even when needed, and will use sensor fusion techniques to manage the possible measurement errors.

3. LOCALIZATION USING RATIO OF SQUARE-DISTANCES

Given three radiation sensors S_1, S_2, S_3 placed on a two dimensional surface, the intensities due to a point radioactive source measured by the sensors are given by

$$I_i^{'} = \frac{I_i - B_i}{E_i \times 2.22 \times 10^6} = \frac{A_0}{(x_s - x_i)^2 + (y_s - y_i)^2},$$

for $i = 1, 2, 3$, according to the sensing model in Section 2.1. The ratio of intensity measurements by two of the sensors, S_i and S_j, due to the radioactive source is given by

$$\frac{I_j^{'}}{I_i^{'}} = \frac{(x_s - x_i)^2 + (y_s - y_i)^2}{(x_s - x_j)^2 + (y_s - y_j)^2},$$

which is equivalent to the ratio of the square-distance between the radioactive source and S_j to the square-distance between the radioactive adiation source and S_i. Using this relationship, we define a locus that contains all the possible points satisfying the ratio of intensity measurements by S_i and S_j as

$$L_{ij} = \left\{ (x, y) \left| \frac{(x - x_i)^2 + (y - y_i)^2}{(x - x_j)^2 + (y - y_j)^2} = \frac{I_j^{'}}{I_i^{'}} \right. \right\}.$$

The *ratio of square-distance* (RoSD) method computes the source position estimate $(\hat{x}_s, \hat{y}_s)$ by solving the equation $L_{12} = L_{13}$, which is equivalent to finding the intersection points of the two loci. The locus L_{ij} is an Apollonius Circle [13] centered at $\left(L_{ij}^x, L_{ij}^y\right)$ and with radius L_{ij}^r, where

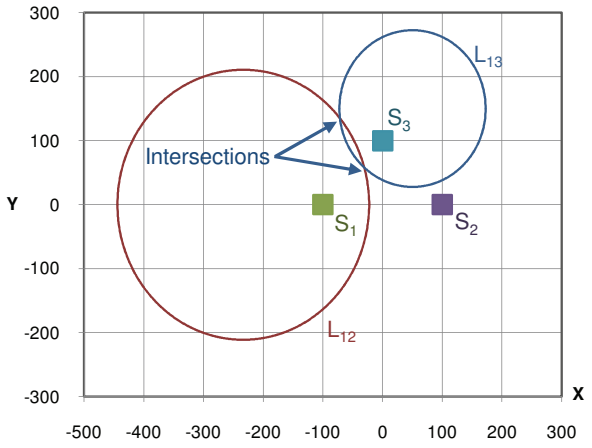


Figure 1: Illustration of loci L_{12} and L_{13}, where $S_1 = (-100, 0)$, $S_2 = (100, 0)$, $S_3 = (0, 100)$, $I_2^{'}/I_1^{'} = 0.40$, and $I_3^{'}/I_1^{'} = 3.00$.

$$L_{ij}^x = \frac{I_i^{'} x_i - I_j^{'} x_j}{I_i^{'} - I_j^{'}} \quad L_{ij}^y = \frac{I_i^{'} y_i - I_j^{'} y_j}{I_i^{'} - I_j^{'}}$$

$$L_{ij}^r = \sqrt{\frac{I_i^{'} I_j^{'}}{\left(I_i^{'} - I_j^{'}\right)^2} \left[(x_i - x_j)^2 + (y_i - y_j)^2\right]},$$

for $I_j^{'}/I_i^{'} \neq 0, 1, \infty$.

Thus, $L_{12} = L_{13}$ may have at most two solutions. An example of the two loci is shown in Figure 1. A closed-form formula that solves the equation returns two solutions as shown in Figure 1. In the case of the two estimates produced, either estimate could be the true position of the source. We do not know which one is the true source position because a strong intensity source located far away from the sensors will produce similar readings as a low intensity source located close to the sensors. One of the two estimates is a false position, and we call this estimate a *phantom estimate*. Identifying the phantom estimate is not straightforward. The fusion algorithm in Section 4 accomplishes this task by using measurements from $n > 3$ sensors. In particular, it exploits the availability of more sensors to produce a robust solution in the presence of noise (i.e., background radiation), statistical errors (due to the random nature of the source being measured), and measurement errors (due to the limited sensor efficiencies).

4. SENSOR DATA FUSION

The ratio of square-distance (RoSD) method with three sensors produces up to two position estimates of the source depending on the placement of the sensors and the location of the source. When there are two estimated positions, it is known that the ambiguity can be resolved by using 4 (instead of 3) sensors [3], *in the case of ideal deployment conditions without noise, statistical errors, and measurement errors*. In a real deployment, however, noise and errors must be considered, in which case a larger number of sensors is needed to produce reliable localization results. Specifically,

if n sensors are used, $\binom{n}{3}$ distinct subsets of 3 sensors are available, each of which will produce up to two source position estimates by RoSD. We will refer to the set of position estimates produced by RoSD as the *candidate estimates* denoted by C. Using C, we will design a fusion algorithm—one that is robust to noise and statistical/measurement errors—to compute a *fused estimate* as the final estimate of the true source position.

A simplistic design of the fusion algorithm is to select the estimate having the smallest maximum distance to all the sensors in the surveillance area. The design gives preference to the candidate estimate that has the highest signal-to-noise (SNR) ratio. The rationale is that, by the inverse-square law, radiation sensors closer to the source will record stronger signals of the source (i.e., higher intensity readings induced by the source) compared with sensors that are farther away from the source. Clearly, the stronger signals are less affected by the noise and hence are more reliable. When phantom estimates are part of the candidate estimates, however, this reasoning may not always hold. This is because a phantom estimate can appear to have a strong signal but is in fact produced as a side effect of RoSD localization and does not well approximate the true source location. As a solution to the problem, the design of a *clustering algorithm* to eliminate the phantom estimates with high probability is the subject of Section 4.2.

4.1 Existing fusion methods

We now review two state-of-the-art fusion algorithms in the literature for radiation localization. Ajith Gunatilaka *et al.* [6] propose the maximum likelihood estimator (MLE) method to estimate the parameters of a radioactive source, i.e., the 2D coordinates of the source location (x, y) and the strength of the source A_0, by finding a solution that best fits the measurement data to the sensing model. Using a radiation intensity model similar to ours, the estimated radiation intensity is calculated. The main idea of MLE is to find values of x, y, and A_0 that minimizes the error between the estimated radiation intensity and the actual measurements recorded by the sensors. This approach does not require a separate fusion step when scaling to $n > 3$ sensors, unlike our algorithm in Section 4.3, which treats the localization and fusion steps as separate problems. In MLE, finding the best-fit solution maps into a multi-dimensional optimization problem for which there are known existing solutions. Specifically, they use the *fminsearch* routine in MATLAB©, which implements the derivative-free downhill simplex method to solve the problem numerically. The disadvantage of the approach is that the result may not be the global optimum when phantom estimates are considered.

Rao *et al.* [16] proposed the mean-of-estimator method (MoE) to fuse the candidate estimates. The MoE method computes the fused estimate as the mean of all the candidate estimates. The advantage of the MoE method is that it has linear time complexity and generally runs significantly faster than MLE. The main drawback of MoE is that it is not explicitly designed to eliminate the phantom estimates in the fusion process. Phantom estimates can be detrimental to the localization accuracy, particularly when they appear to be produced by strong (and hence presumably reliable) sensor readings. As a result, their algorithm can produce large localization errors when a significant fraction of the candidate estimates are phantom estimates.

4.2 Data fusion with phantom estimates eliminated by clustering

Let us take a look at the characteristics of candidate estimates produced by RoSD. Assume an ideal deployment situation in which there is no noise due to background radiation, and the sensor measurements are perfectly accurate. As discussed in Section 3, RoSD may produce one or two candidate estimates with measurements from a group of three sensors, S_1, S_2, and S_3. In the case that there are two estimates, the phantom estimate is the estimate that is farther from the true source position compared with the other estimate. Note that phantom estimates are produced not because the localization algorithm is not accurate, but because the localization algorithm does not have sufficient information to compute an unequivocal solution. Suppose now that an additional sensor S_4 is available, and consider the group of three sensors S_1, S_2, and S_4. One or two candidate estimates are again produced. In the assumed ideal deployment situation, one of the two estimates in each group (i.e., the group S_1, S_2, S_3 and the group S_1, S_2, S_4) will coincide at the true position of the source. The other estimates, namely the phantom estimates, from the two groups are highly unlikely to coincide because two of the sensors from the groups are located differently. Thus, by counting the number of estimates at different positions, the position with the highest frequency would be the true position of the source.

In general, with n sensors and assuming the sensor placements are not degenerate, there are $\binom{n}{3} = \frac{n}{6}(n-1)(n-2)$ groups of sensors available to perform RoSD localization, and each group produces a set of candidate estimates C_i for $i = 1, 2, \ldots, \frac{n}{6}(n-1)(n-2)$. At least one of the estimates in C_i is the true position of the source; thus there are at most $\binom{n}{3}$ estimates located at the true position of the source, and at most $\binom{n}{3}$ phantom estimates. The source position can be estimated by $(\hat{x}_s, \hat{y}_s) = \bigcap C_i$ for $n > 3$. With precise intensity measurements and lack of noise from background radiation, $n = 4$ is guaranteed to accurately estimate the position of the source. Generalizing to many groups of three sensors, we see that in an ideal deployment situation, one of the estimates from *every* group will coincide at the true source location, thus solving the localization problem redundantly.

When there are noise and statistical/measurement errors, the noise/errors will scatter the non-phantom estimates and hence they will not exactly coincide. The extent of the scattering depends on the noise/error magnitudes. As a result, a simple intersection of the candidate estimates will likely produce a null set. Nevertheless, in spite of the scattering, the candidate estimates that are not phantom estimates will likely be close to each other and concentrate around the true position of the source. On the other hand, the candidate estimates that are phantom estimates will not cluster together. Instead of finding the intersection of the candidate estimates, we can therefore "zoom in" to the true source position by identifying a most dense cluster of the estimates. Ideally, the most dense cluster found does not contain any phantom estimate, so that an unbiased estimator of the candidate estimates in the cluster will give a fused estimate that closely approximates the true source location. In practice, the phantom estimates may not be eliminated completely, but if most of them are eliminated, their total contribution to the fused estimate will be small.

In summary, we divide the localization problem using n sensors, under realistic noise and error conditions, into two solution steps:

1. **Subproblem P1:** *Clustering.* Find a smallest region in the surveillance area that contains most, if not all, candidate estimates that are not phantom estimates.

2. **Subproblem P2:** *Fusion.* Compute the fused estimate as an unbiased estimator of all the candidate estimates in the cluster found above.

4.3 Iterative pruning (ITP) clustering algorithm

There are existing clustering algorithms that are relevant to the first solution step above. Algorithms such as k-mean [7], CURE [5], and DBSCAN [4] are not directly applicable because they are concerned with classifying all samples into a number of closest clusters, whereas our problem aims to find one largest cluster that will likely contain most of the non-phantom estimates but few of the phantom estimates. Another algorithm, the Quality Threshold (QT) clustering algorithm [9], was invented by Heyer to cluster gene expression patterns. In their domain, the algorithm groups genes of high similarity into the same cluster. The measure of similarity is user defined, although the correlation of the gene expression is often used in their problem. The quality of the cluster computed is ensured by specifying a threshold such that all genes with a similarity measure within the threshold will fall into the same cluster. Consider that there are n genes. The algorithm first builds n candidate clusters. It then outputs the largest cluster among the n candidates as a result. The genes produced are removed from the pool of n genes and the algorithm repeats the same step for the remaining genes. The candidate cluster is built by selecting the i-th gene from the pool, and then iteratively including the most similar gene, the second most similar gene, and so on, until the similarity threshold is exceeded.

In our problem domain, the genes can be interpreted as our candidate estimates, and the similarity measure corresponds to the distance between the candidate estimates. The QT algorithm will then basically associate each candidate estimate (the gene) with a cluster such that the diameter of the cluster does not exceed d (the similarity threshold). The difference is, we are only interested in the maximum size cluster in our problem, and therefore do not need to iteratively find the next largest cluster as in the original algorithm.

While QT is useful for our purpose, the main disadvantage of QT clustering is that the time complexity is very high, $O\left(|\mathbf{C}|^4\right)$, even for finding the maximum size cluster only. For our problem, the algorithm will not scale to a large-size network (i.e., the number of sensors n is large and there are $O(n^3)$ samples for clustering) required for, say, monitoring of large city areas. We now propose an algorithm, which we call the *iterative pruning* (ITP) algorithm, that for our problem, can achieve similar performance as QT clustering, but has a greatly reduced time complexity. Specifically, ITP has a worst case time complexity of $O\left(|\mathbf{C}| \log_2 A\right)$, where A denotes the area of the surveillance region.

The ITP algorithm has two steps. First, it solves Subproblem P1 heuristically by pruning the space (in the surveillance area) with low density of candidate estimates, so that a cluster with a high density of the estimates remains. Second, it solves Subproblem P2 by computing a weighted centroid of the candidate estimates in the cluster as the fused estimate. The pseudo-code of the ITP algorithm is given in Algorithm **1**. The algorithm takes three input parameters. The $\mathbf{C}$ parameter takes the union of all the candidate estimates $\mathbf{C} = \bigcup C_i$ produced by the RoSD algorithm. The N and d parameters limit the maximum number of estimates in the remaining region, and the maximum size (in terms of area) of the smallest region, respectively.

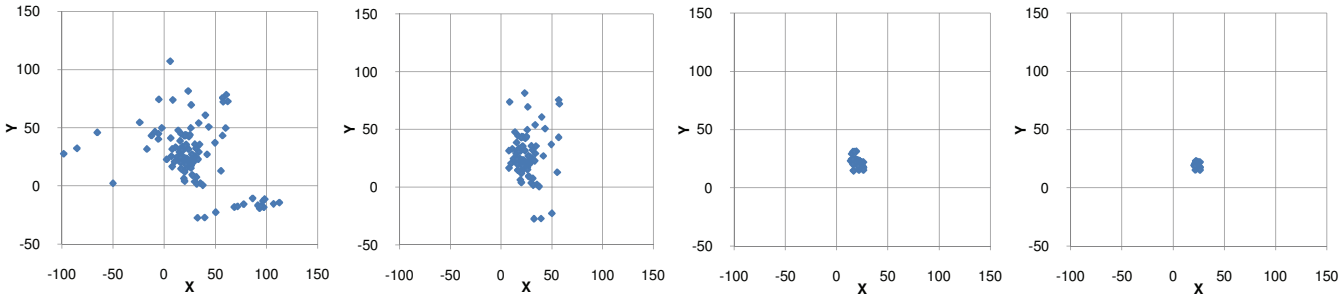
The algorithm begins with the smallest rectangle that bounds all the candidate estimates is determined. The bounding rectangle is then divided into 5 regions $r_a = r_0 \cup r_1 \cup r_4 \cup r_5$, $r_b = r_2 \cup r_3 \cup r_6 \cup r_7$, $r_c = r_0 \cup r_2 \cup r_4 \cup r_6$, $r_d = r_1 \cup r_3 \cup r_5 \cup r_7$, $r_e = r_4 \cup r_5 \cup r_6 \cup r_7$, and 8 subregions $r_0, r_1, \ldots, r_7$. The subregion labels are shown in Figure 3. The number of candidate estimates in each of the sub-regions inside the rectangle is tallied, and then the region containing the most estimates is selected for the next iteration.

The algorithm continues until the number of candidate estimates remaining is less than N and the area of the bounding rectangle is smaller than $d \times d$. In each iteration, the size of the bounding rectangle is reduced by at least half. This guarantees that the algorithm will terminate in $O\left(\log_2 A\right)$ iterations. Unlike existing partitioning algorithms such as the generalized bisection method in [19] which iteratively divides the search space in halves, ITP divides the space into five overlapping regions. The overlapping minimizes the chance that the algorithm will incorrectly prune a region because a cluster is concentrated near the division boundaries. For instance, Figure 4 shows two scenarios where non-overlapping partitioning will fail. In particular, Figure 4(a) shows a scenario in which the candidate estimates are concentrated at the center of r_c. If a non-overlapping region division is used, the algorithm would choose either $r_0 \cup r_4$ or $r_2 \cup r_6$. Since the numbers of candidate estimates in both regions are close, however, it is better to choose $r_0 \cup r_2 \cup r_4 \cup r_6$ for the next iteration.

In the second part of the algorithm (Lines 32 to 37), the weighted center of the cluster is computed as a solution to Subproblem P2. The weighting gives estimates produced by sensors close to the source higher weights because their measurements are less influenced by the background noise. Although this is similar in concept to the use of SNR in wireless communication, there are subtle differences. In wireless communication, the SNR provides a measure of how intrusive the background noise is compared with the goodness of the signal received. In particular, the SNR quantifies how likely the data received will be correctly decoded in spite of the noise. These quantities can be directly measured at the receiver side. In our problem, the SNR of a candidate estimate is a function of the SNRs of the three measurements that produced the candidate estimate. Essentially, the SNR of a candidate estimate measures the goodness of the estimate given the goodness of the measurements by the individual sensors. Without knowledge of the sensor location, using the average of the SNRs of all the three measurements gives us an unbiased estimate of the SNR of the candidate estimate.We define the SNR of a candidate estimate as

$$\mathrm{SNR}\left(c\right) = \frac{\sum_{x=\{i,j,k\}} I_x}{\sum_{x=\{i,j,k\}} B_x}$$

where S_i, S_j, S_k are the three sensors that produce the estimate c.



(a) All candidate estimates (b) Iteration 2 (c) Iteration 6 (d) Final Result

Figure 2: The candidate estimates are pruned in each iteration until the remaining estimates are clustered within a $d \times d$ rectangular region.

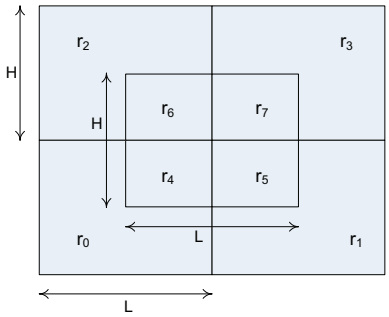
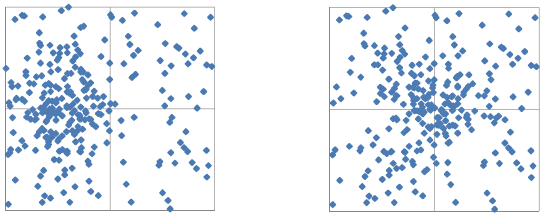


Figure 3: Region division in ITP algorithm.



(a) Cluster located at the division of two regions.

(b) Cluster located at the center, division of four regions.

Figure 4: Two examples illustrating that non-overlapping region division may lose accuracy when the cluster is located near the division of regions.

Figure 2 shows a sample output tracing selected iterations of the ITP algorithm. The candidate estimates fed into the algorithm (Figure 2(a)) are iteratively pruned such that the region having the largest number of estimates remains after each iteration. The algorithm stops when the number of estimates remaining is less than N and the area of the region is less than $d \times d$. The weighted center of the remaining estimates is computed as the fused estimate. As the number of sensors increases, the candidate estimates produced increases as $O\left(n^3\right)$. As a result, the algorithm can identify a small region that will most likely contain the true source location as the region with a high density of the estimates. The experimental results in Section 5.5 confirm that the accuracy of ITP increases with the number of sensors used.

5. TESTBED AND SIMULATION RESULTS FOR 0.911 μCI SOURCE

In this section, we evaluate the performance of our localization algorithm for an extremely low-intensity (0.911 μCi) radiation point source using both testbed and simulation experiments. Section 5.1 describes the setup of the simulation experiments. Section 5.2 presents the calibration procedure to obtain the efficiency coefficients of sensors described in Section 2.1. In Section 5.3, we validate the radiation sensing model and our simulation model to ensure that the simulation results are realistic. After that, in Section 5.4, we characterize the performance of the RoSD localization algorithm with three sensors and identify the source of localization errors in the algorithm. We also compare the performance of RoSD with the log-space DTOA localization algorithm in [15]. Lastly, in Section 5.5, we evaluate the performance of the ITP fusion algorithm, and compare its performance with that of the MoE and MLE algorithms in the case of more than 3 sensors.

To ensure that the results are comparable, we will implement the same scenario for both the simulation and testbed experiments. Specifically, we will use the following small-scale setup that is commensurate with the extremely low intensity radioactive source we use in this section. There are 18 sensors in a 50 cm $\times$ 50 cm surveillance region; their positions are given in Table 1. Not all the sensors will be activated in all the experiments. A CS-137 radiation point source (which emits Gamma rays) of intensity $A_0 = 0.911$ μCi is located at $(x_s, y_s) = (19.09, 19.09)$. All the distance measurements are in centimeters (cm). The placement scenario is illustrated in Figure 5.

For the testbed experiments, the distances between the radioactive source and the sensors were calculated according to Table 1. The radioactive source was then placed at the calculated distances from the sensors. Traces for S_1 to S_{18} were collected by polling the corresponding sensors every 4 seconds. A total of 3738 measurements were collected for each sensor. Note that the traces are not synchronized because they may not have been taken at the same time. However, this does not affect the experimental results because the underlying nuclear disintegrations follow a memoryless Poisson process. The traces collected were then used to drive the experiments.

Note that the radioactive source used in the testbed experiments needs to be of extremely low intensity in order to be safe for experimentation. The intensity is so low that the

Algorithm 1 ITP($\mathbf{C}$, N, d)

r_S = smallest rectangular region that bound all estimates $c \in \mathbf{C}$.
Let b and t denotes the bottom-left and top-right coordinate of r_S.
Divide the rectangle into 8 regions r_i for $i = 0, 1, \ldots, 7$ as in Figure 3.
Let k_w denotes the number of estimates in region w.
Initializes $k_w = 0$, for $w = r_0, r_1, \ldots, r_7$.
for all $c \in \mathbf{C}$ **do**
  for $i = 0$ to 7 **do**
    if c is inside r_i **then**
      $k_{r_i} = k_{r_i} + 1$
    end if
  end for
end for
Let $r_a = r_0 \cup r_1 \cup r_4 \cup r_5$; Let $r_b = r_2 \cup r_3 \cup r_6 \cup r_7$
Let $r_c = r_0 \cup r_2 \cup r_4 \cup r_6$; Let $r_d = r_1 \cup r_3 \cup r_5 \cup r_7$
Let $r_e = r_4 \cup r_5 \cup r_6 \cup r_7$
$r_S = \arg\max_{w \in \{r_a, r_b, r_c, r_d, r_e\}} k_w$
if $r_S = r_a$ **then**
  $t.y = 0.5\,(t.y + b.y)$
else if $r_S = r_b$ **then**
  $b.y = 0.5\,(t.y + b.y)$
else if $r_S = r_c$ **then**
  $t.x = 0.5\,(t.x + b.x)$
else if $r_S = r_d$ **then**
  $b.x = 0.5\,(t.x + b.x)$
else
  $b.x = 0.25\,(t.x + b.x)$; $b.y = 0.25\,(t.y + b.y)$
  $t.x = 0.75\,(t.x + b.x)$; $t.y = 0.75\,(t.y + b.y)$
end if
if $|r_S| > N$ or $(t.x - b.x)\,(t.y - b.y) > d$ **then**
  return ITP_Fusion($\forall c \in r_S$, N, d)
else
  $e = (0, 0)$, $s = 0$
  for all $c \in r_S$ **do**
    $e.x = e.x + c.x \times \mathrm{SNR}\,(c)$; $e.y = e.y + c.y \times \mathrm{SNR}\,(c)$
    $s = s + \mathrm{SNR}\,(c)$
  end for
  return $\left(\frac{e.x}{s}, \frac{e.y}{s}\right)$
end if

sensors observe only the background radiation when they are placed more than 4 feet away from the source. This caused us to use a small size region in the experiments.

5.1 Simulation model

We have also created a simulation model for the above testbed scenario to allow us to systematically explore the parameter space that we cannot do in actual experiments. The simulation results will, for example, reveal the sensitivity of the RoSD and ITP algorithms to varying source strengths. Radiation counts emitted by the source are modeled as a Poisson process. Further, we evaluate our algorithm using three variations of the radiation sensors:

1. **Real sensor** (testbed): We use the RFTrax RAD-CZT radiation sensor [1] as the real sensor in our testbed experiments. The RFTrax sensor uses the Cadmium Zinc Telluride (CZT) detector technology and has 0.1 mREM/hour sensitivity.

2. **Ideal sensor** (simulations): We simulate an ideal sensor as one that is capable of detecting all gamma rays radiated, i.e., its sensitivity is $E_i = 1.0$ CPM/DPM so that its measurements are totally accurate.

Table 1: Sensor coordinates used in 0.911 μCi source experiments.

Sensor	X (cm)	Y (cm)	Sensor	X (cm)	Y (cm)
S_1	36.52	29.89	S_{10}	38.19	16.33
S_2	27.45	29.70	S_{11}	4.43	0.47
S_3	26.13	40.46	S_{12}	14.65	5.60
S_4	18.53	37.08	S_{13}	19.46	7.10
S_5	10.11	41.57	S_{14}	4.64	35.28
S_6	32.31	12.00	S_{15}	36.15	0.00
S_7	0.00	0.00	S_{16}	23.65	7.67
S_8	5.24	21.10	S_{17}	0.81	7.77
S_9	3.13	15.77	S_{18}	42.68	24.01

Figure 5: Placement of radioactive source and radiation sensors in experiments.

3. **Model of real sensor** (simulations): We create a simulation model of the RFTrax RAD-CZT radiation sensor using specifications from the manufacturer. We discussed the architecture of the sensor with a knowledgeable Engineering Manager in RFTrax, Inc. We obtained needed information to model the firmware processing in the RFTrax sensor accurately. Essentially, there are some major differences between the RFTrax sensor and a model of the ideal sensor: (1) The efficiency coefficient of RFTrax sensor is much less than 1.0 CPM/DPM, but is instead given in Table 2; (2) The RFTrax sensor keeps track of the radiation count in the past second only, instead of for minutes. Hence, a CPM count reported by the sensor is the number of disintegrations detected in the past second multiplied by 60. The measurements are therefore inaccurate, particularly for the extremely low intensity 0.911 μCi CS-137 source used in the experiments.

5.2 Testbed sensor calibration

As discussed in Section 2.1, the sensor calibration procedure determines the sensor efficiency and the background radiation intensity at the sensor location. During the calibration process, three RFTrax RAD-CZT radiation sensors were placed on a table, without the presence of a radioactive source, to collect readings of the background radiation intensity. Then, we used a CS-137 radioactive source of $A_0 = 0.911$ μCi to collect another set of readings due to the source intensity. The calibration results are shown in Table 2.

5.3 Validation of sensing model

We have validated the sensing model in Section 2.1 by collecting traces of CPM counts obtained with the 0.911 μCi

Table 2: Efficiency coefficients of three real RFTrax sensors, average measured intensities of CS-137 radioactive source, and measured background intensities.

Sensor Serial#	Avg Source (CPM)	Avg Background (CPM)	Efficiency (CPM/DPM)
000877	4202.6859	10.5217	0.2078%
000809	6274.8549	13.6521	0.3103%
000841	4882.6332	18.4220	0.2414%

Table 3: Comparison between real measurements from three RFTrax sensors and the radiation intensity model.

Sensor Serial#	Distance (cm)	Avg Source (CPM)	Model (CPM)	Difference (CPM)
	13.5	37.0099	48.0821	11.0722
000877	24.2	21.1236	24.3667	3.2431
(S_1, S_4,	14.0	32.3821	45.6667	13.2846
S_7, S_{10},	23.7	19.9144	24.8235	4.9091
S_{13}, S_{16})	21.7	22.1469	26.9776	4.8307
	21.5	23.1495	27.2267	4.0772
	22.5	25.3450	28.0667	2.7218
000809	15.0	38.3307	40.1226	1.7920
(S_2, S_5,	16.3	32.8900	36.7992	3.9093
S_8, S_{11},	14.2	40.3292	42.6366	2.3075
S_{14}, S_{17})	25.6	24.3019	25.8723	1.5704
	24.1	25.8622	26.8286	0.9665
	20.5	19.7025	20.5222	0.8197
000841	18.0	21.0433	23.4930	2.4496
(S_3, S_6,	27.0	15.3083	16.2867	0.9784
S_9, S_{12},	19.3	19.5326	21.8044	2.2718
S_{15}, S_{18})	12.0	35.0893	39.7071	4.6178
	12.3	34.5478	38.3007	3.7529

CS-137 radioactive source placed at specified distances from the sensors. The measured radiation intensities and the corresponding distances between the sensor and the source were recorded. The experiment was repeated 6 times with the source located at a different distances to the sensors. A total of 18 traces were collected. The average intensity computed from the traces is compared with the value computed from the sensing model. The results are shown in Table 3. They show that the the model results and actual measurements differ by less than 5 CPM for 89% of the time.

5.4 Performance of RoSD

5.4.1 Evaluation methodology

In this section, we evaluate the performance of the RoSD algorithm in both testbed and simulation experiments. We activate sensors S_7, S_{15}, and S_4 and measure the performance of RoSD only, i.e., without running the ITP fusion algorithm. Because the fusion algorithm is not used, multiple candidate estimates produced by RoSD are resolved using the smallest maximum distance approach, i.e., we select the estimate that has the higher SNR. To ensure that the simplified algorithm will work, the 3 sensors are placed such that the radioactive source is contained within the triangle formed by the sensors.

During the simulation experiments, the sensors sample the radiation intensity m times, and the average intensity is computed. The radiation intensity detected by the sensor is modeled as a Poisson variable with $\lambda = I_i$. The values of $I_{15}^{'}/I_7^{'}$ and $I_4^{'}/I_7^{'}$ are then calculated using the average intensities measured by each sensor. We then invoke the RoSD algorithm to compute a position estimate. Each simulation run is repeated 1000 times for each w value of $1, 2, \ldots, 100$.

5.4.2 Localization error

The impact of the probabilistic radioactivity on the measured radiation intensities can be seen in Figure 6(a). In this experiment, the model of ideal sensors is used; hence, the only source of variability in the system is due to the underlying Poisson process of radioactivity. Given the standard error of the measurements,

$$\sigma_{\hat{I_i^{'}}} = \frac{\sigma_{I_i^{'}}}{\sqrt{m}} = \frac{\mu_{I_i^{'}}}{\sqrt{m}},$$

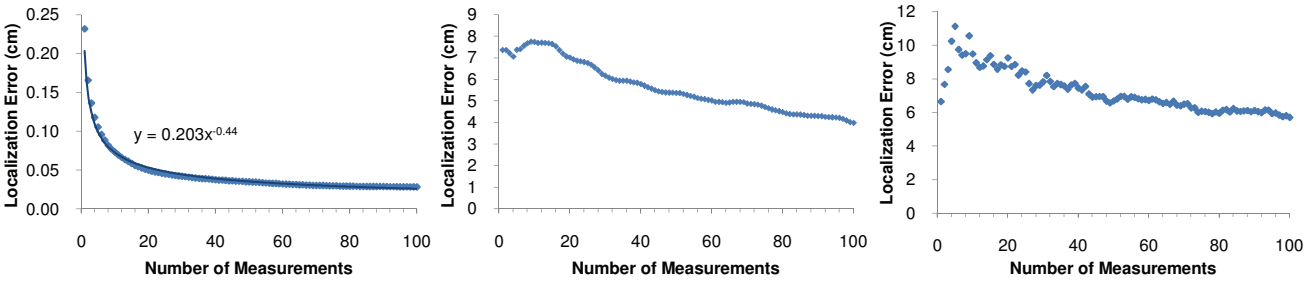
we expect that the localization error decreases like $O\left(\frac{1}{\sqrt{m}}\right)$. In the simulation results, the localization errors as a function of number of measurements are best fitted to the curve $y = 0.203m^{0.44}$, which is close to the expected $O\left(\frac{1}{\sqrt{m}}\right)$. Moreover, the localization error is small—an average error of 0.05 cm with 20 measurements in a testbed where the maximum distance between two sensors is 41.45 cm.

We now introduce measurement errors in the experiment. To do that, we repeat the above experiment, but now with a model of the RFTrax sensor. In this case, the localization error increases by 125× on average, as shown in Figure 6(b). This increase in error is solely due to the limitation of the actual sensor that the particle count during the last second only is remembered. Should the sensor record a longer history of the counts, the localization error would have been lower. This result shows that limitations in specific sensor implementations can contribute significant errors in RoSD localization.

We now report results for the actual testbed experiments. In this case, probabilistic background radiation is also present besides the measurement errors. Because of the background radiation, the localization error increases by 29% compared with simulations using the RFTrax sensor model. In addition, there are high fluctuations in the localization error when the number of measurements is small. This is shown in Figure 6(c), and suggests that a small number of measurements is not sufficient for accurately localizing a source, given the high variance of the underlying Poisson process. The problem is similarly amplified in the case of the real RFTrax sensors, due to the short, one-second measurement window remembered by the sensors. Consequently, averaging over a small number of measurements does not give accurate estimates of the true source intensity, which is needed for RoSD to work well.

5.4.3 Comparison of RoSD with log-space DTOA

For comparison with the RoSD algorithm, we have implemented the DTOA algorithm in [22] and adapted it to log space as described in [15] for radiation localization. Our performance comparison is therefore between RoSD and the log-space DTOA algorithm. In this experiment, all distances are given in a generic distance unit since the exact physical unit does not matter. We construct a simulation scenario in which three sensors are placed at coordinate $(0, 0)$, $(1000, 0)$ and $(500, 1000)$. We systematically set the radioactive source location to be at each position (in steps of 2 dis-



(a) Simulated ideal sensors (b) Simulated non-ideal sensors (c) Real sensors

Figure 6: Localization error with respect to the number of measurements using 3 sensors. The source is positioned such that ambiguity can be resolved using mini-max distance approach.

tance units) in $(x_s, y_s) = ([-140, 1140], [-300, 1000])$. For each position, we calculate the ratio of intensities by two sensors and feed it into the RoSD algorithm. Similarly, the logarithm of intensity differences is calculated, and we feed it into the log-space DTOA algorithm. All inputs are deterministic, and noise and errors are not present. The log-space DTOA algorithm is an iterative algorithm. It terminates when the error in the distance ratio is smaller than ϵ. We set the value of ϵ to be 1.0×10^{-12}, in which case the algorithm usually terminates in less than 100 iterations. Nevertheless, we set the maximum number of iterations allowed to 10,000, to ensure that any error in the position estimate is not due to premature termination of the algorithm.

Both the RoSD and log-space DTOA algorithms may produce up to two estimates for each set of measurements from 3 sensors. In the case that there are two estimates, both algorithms select the estimate that has the smaller maximum distance to all the sensors. Hence, both algorithms are not able to localize a source that is not contained by the triangle formed by the sensors. By our calculations, approximately 21.46% of the simulated source positions are not contained within this triangle. This is the lower bound of the errors, i.e., both algorithm will not able to produce correct estimates for at least 21.46%. We compute the distance between the estimate and the true source position, and conclude that the algorithm successfully localizes the source if the localization error is less than 2 distance units. Otherwise, we conclude that the algorithm produces a wrong estimate. If the algorithm does not produce any estimate, we count that as no estimate. Figure 7 shows the error rates for the RoSD and log-space DTOA algorithms. In general, with 3 sensors, RoSD algorithm improves 6% over log-space DTOA algorithm. The log-space DTOA algorithm produces 7.8× more no estimates than RoSD, but RoSD produces 6% more wrong estimates than log-space DTOA. In total, RoSD and log-space DTOA do not produce a correct estimate for 26.70% and 28.35%, of the cases, respectively. These numbers are 5.24% and 6.89% above the 21.46% error lower bound, respectively, showing that the performance loss is minor. With one additional sensor, RoSD with the ITP fusion algorithm will produce correct estimates virtually all of the time.

5.4.4 *Run time of algorithms*

In terms of execution time, RoSD runs 497 times faster than log-space DTOA in our experiments, as shown in Table 4. The reported execution times are for the algorithms

Table 4: Execution time of RoSD and Log-space DTOA.

Algorithm	RoSD	Log-space DTOA
Execution Time (s)	0.395	196.413

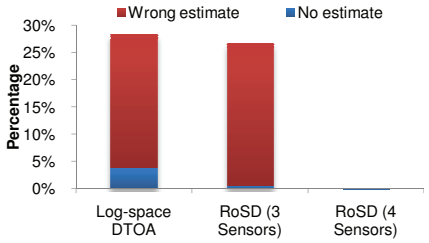


Figure 7: Error rates comparison between RoSD and Log-space DTOA.

to localize all the source positions in the above experiment setup. Our execution time results were taken on an Intel Pentium 4 HT 3.40 GHz machine with 2 GB RAM. The results were obtained using the hardware performance counter in the CPU, which has high resolution and is minimally influenced by software overheads.

5.5 Performance of ITP fusion algorithm

5.5.1 *Evaluation methodology*

We now evaluate the performance of the sensor data fusion algorithm. We repeat the experiments in the previous section, but now with more than three sensors and with fusion of the candidate estimates by the ITP algorithm. In the experiments, we systematically activate sensors S_1 to S_n for $n = 4, \ldots, 18$. We compute the union of all the estimates and feed it into the ITP algorithm. ITP outputs the fused estimate, which is compared against the true source position. The localization errors produced by ITP are compared against those of the Maximum Likelihood Estimator (MLE) and mean-of-estimator (MoE) methods in Section 4.1. For clustering, comparison results with QT clustering are also shown. One issue is that the MoE algorithm assumes that the localization algorithm produces only one candidate estimate per run. In our case, the localization algorithm may produce up to two estimates for each subset of 3 sensors. Because selecting between the two candidate estimates for use in MoE has a large performance impact, we evaluate two possible approaches:

- **Smaller maximum distance to sensors:** Labeled as *MoE* in the reported results (Figure 8), this approach selects the estimate that has the smaller maximum distance between the estimate and all the sensors.
- **Random:** Labeled as *MoE/R* in the reported results, this approach randomly selects one of the two candidate estimates for use with MoE.

5.5.2 Comparison of localization errors

Two major factors determine the errors of fused estimates. The first factor is the error produced by the localization algorithm. The second factor is the distance between the two candidate estimates. The former is determined by the accuracy of the localization algorithm itself and the magnitude of noise that gets included in the computation. The noise is in turn affected by factors such as the sensor placements and the accuracies of the measurements themselves. For example, although MoE has no built-in design to disambiguate each pair of estimates, the omission will not have a large performance impact if the distance between the two candidate estimates is small, i.e., selecting one versus the other will not matter much. ITP, in contrast, has a built-in clustering step to disambiguate the candidate estimates. This is very useful in general, but will appear less so if the phantom estimates are close true position of the source.

The above observations are illustrated in Figure 8(a) and 8(b). With ideal sensors, RoSD produces virtually zero localization errors. Hence, the distances between candidate estimates are huge in comparison with the (almost zero) localization errors. In this case, MoE and MoE/R produce large localization errors, but ITP produces practically no error because the algorithm accurately disambiguates the two candidate estimates. With non-ideal sensors, the localization error is much larger. Errors due to the probabilistic measurements cause the estimate to deviate by as large as the distance between two sensors. At the same time, the distance between candidate estimates is roughly the same as the distance between the source and each estimate. In this situation, averaging over all the estimates performs as good as the more expensive ITP algorithm. The performance advantages of ITP become more clear when background radioactivity is considered as shown in Figure 8(c). With radiation sensors of higher efficiency, the ITP algorithm will perform better as the localization error is reduced.

The parameter d in ITP controls the maximum size of the region that will not contain the phantom estimates. With error-free measurements, small value of d is sufficient. When considering the measurement errors, a larger value of d is required to tolerate the measurement errors such that the algorithm will include most, if not all, non-phantom estimates. Setting the value of the d too large degrades the performance of the algorithm because the algorithm will be more likely to include phantom estimates into the region. In fact, setting $d = \infty$ causes the algorithm to behave similarly to MoE. Implicitly, the value of d gives a boundary to the region where the radioactive source is most likely to contain. The results in Figure 9 show that accurately setting the value of parameter d is not required and a rough value will work well in practice.

5.5.3 Comparison of execution time

In terms of execution time, the results in Figure 10 show that the efficiency of ITP scales well with a larger number of sensors, compared with the QT clustering algorithm and MLE. The reported results measure the average execution time of the ITP, MLE and QT clustering algorithms in 1000 runs. In each run, the execution time of ITP and QT was measured using the CPU performance counter on an Intel Pentium 4 HT 3.40 GHz machine with 2 GB RAM. The execution time of MLE was measured on the same hardware platform using the `profile on -timer cpu` command in MATLAB© because the program was implemented on MATLAB©. The ITP algorithm completes in split seconds in all the runs. Even when the experiment is scaled to 100 sensors, the ITP algorithm takes only about 0.2 seconds to run. For QT clustering we only present data for 4 to 12 sensors because the running time of QT clustering is prohibitive for a larger number of sensors. The execution time of QT clustering will not allow near real-time localization of the source. The execution time of MLE does not show a clear trend with respect to the number of sensors, because the algorithm is sensitive to the starting point of the search. Among all the experiments we ran, the execution time of MLE is at least 252 times of ITP, and on average the execution time is 16000 times longer. The performance of the downhill simplex optimization used in MLE depends on the initial values of the estimated parameters. The algorithm can take many iterations to converge, and frequently exceeded the maximum number of iterations allowed (set to 10,000) in our experiments, if the initial estimates are not close to the actual solution (the true position of the source). Although reducing the maximum number of iterations allowed would bound the execution time, the localization error would increase.

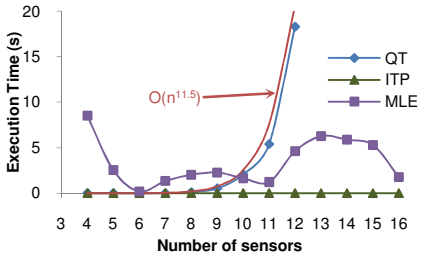
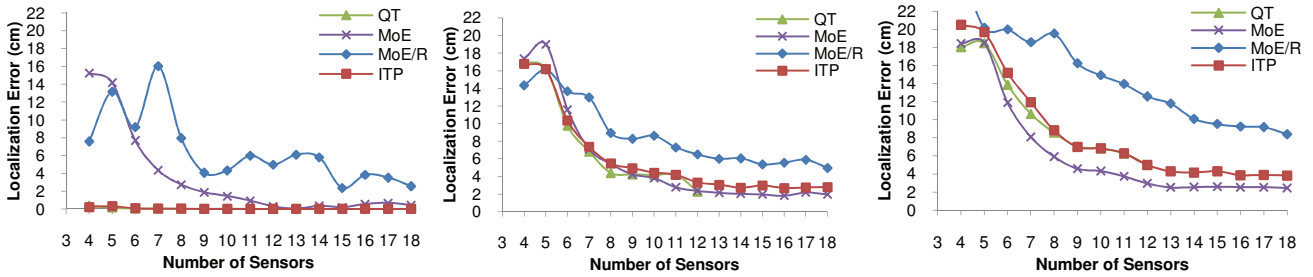


Figure 10: Execution time of ITP, MLE, and QT algorithms.

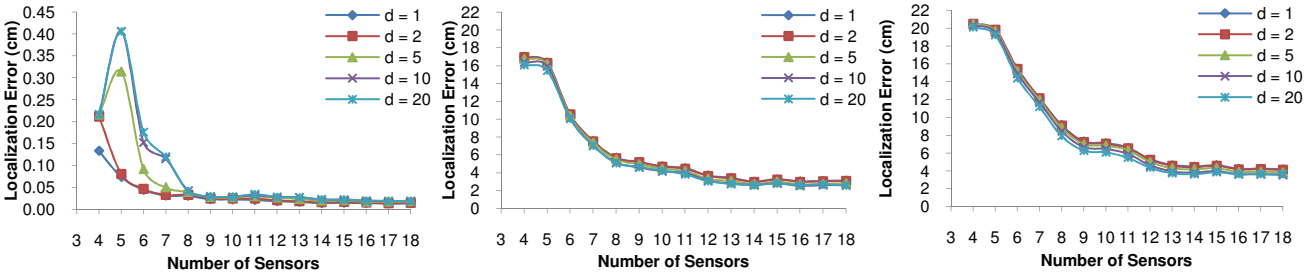
6. SIMULATION RESULTS FOR DIRTY BOMB SCENARIOS

We conducted experiments based on the sensing model in Section 2.1 to evaluate the practicality and effectiveness of ITP localization in real-life situations. We use simulations because actual experimentation with source intensities similar to that of a concealed dirty bomb is not safe. In the simulation setup, up to 100 sensors are located uniformly in a grid, in a surveillance area of size 10×10 m^2, 50×50 m^2, 100×100 m^2, and 200×200 m^2. We model a radioactive source of four different strengths, namely 400 μCi, 4 mCi, 40 mCi, and 400 mCi. The source is uniformly randomly placed in the surveillance area. We compare results by ITP with those of the MoE and MLE algorithms, under the same simulation scenarios. All simulations are repeated at least 6000 times to ensure the statistical significance of the results.



(a) Simulated ideal sensors (b) Simulated non-ideal sensors (c) Real sensors

Figure 8: Localization error of four different variants of QT clustering algorithms, with respect to the number of sensors. Number of measurements and clustering diameter were $m = 20$ and $d = 5$ respectively.



(a) Simulated ideal sensors (b) Simulated non-ideal sensors (c) Real sensors

Figure 9: Localization error with respect to the number of sensors. RoSD was run with $m = 20$ measurements.

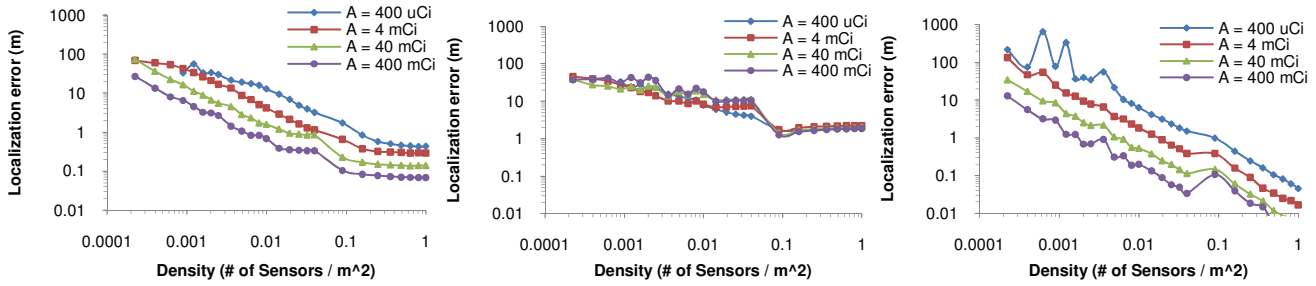
The simulation results show that ITP performs better than both MoE and MLE in localizing a low-dose radioactive source. With one sensor per 33×33 m^2 area or roughly 110×110 ft^2, ITP is able to localize a low level radioactive source of strength 400 μCi with an accuracy of 32.5 m using $m = 20$ measurement samples, whereas MoE and MLE achieve accuracy of 33.6 m and 218.219 m respectively. Figure 11 shows that ITP achieves higher performance gains compared with MoE when the sensor density increases, although MoE has smaller localization errors when the sensor density is very low. MoE outperforms ITP at low sensor densities because the majority of the sensors in this case are not able to detect the source, since they are sparsely located. This leads to a lower number of candidate estimates available. As discussed in Section 4.3, a low number of candidate estimates hurts the performance of the ITP algorithm. When a sufficient number of sensor groups produce estimates, however, the localization errors of ITP dramatically decrease. When the radioactive source has high intensity, MLE outperforms ITP but the MLE execution time is significantly higher and has high variance.

Given that the sensor intensity measurements follow a Poisson process, the variance of the intensity readings is equal to the mean. This indicates that a stronger source will exhibit a larger variance of the measurements. In spite of a large variance, the localization error of the ITP algorithm decreases roughly like $O\left(\frac{1}{\sqrt{A_0}}\right)$, as shown in Figure 12(a), implying that a higher intensity source can be located more accurately. This is a desirable property because high intensity radioactive sources are dangerous, and there is a need to localize such a source quickly and accurately. On the other hand, a stronger source may in fact reduce the localization accuracy of MoE. This is because MoE averages over all the estimates produced, and a larger variance implies a larger error in the estimates produced, for the same number of measurements used. This phenomenon is observed in Figure 12(b), where the localization error increases gradually with the source strength.

Using a larger number of measurements increases the sensing time. A longer sensing time generally decreases the localization error, although the gain in performance is sub-linear. Figure 13 shows the normalized 90-percentile localization error for different numbers of measurements m, for different source strengths. For each source strength, the localization error is normalized to the localization error with $m = 20$. For instance, with a 400 μCi radioactive source, increasing the number of measurements by 250% reduces the localization error by 29% only. The MLE algorithm shows larger improvements when the number measurements increases, when compared with ITP. On the other hand, MoE benefits little when the number of measurements increases. In fact, in some cases the localization error increases when the number of measurements increases.

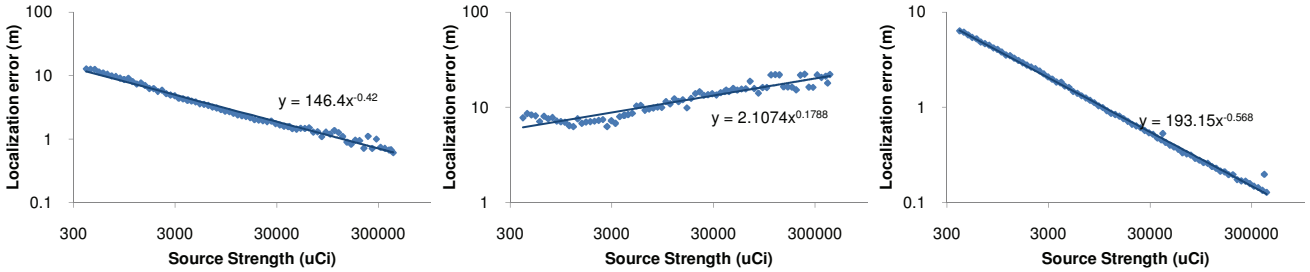
7. CONCLUSION

We have addressed the problem of accurately and quickly localizing a low-level point radioactive source using n sensors, under realistic noise and measurement errors. We have presented an efficient iterative pruning (ITP) algorithm to efficiently fuse the position estimates by groups of 3 sensors to produce a final estimate that is close to the true source position. We have compared the performance of ITP with that of the existing MoE and MLE algorithms. Testbed results and simulation results using a validated simulation model illustrate the performance of our algorithm, including the performance impact of different sources of errors.



(a) ITP fusion algorithm. $d = 7.5$ feet. (b) MoE (c) MLE

Figure 11: Comparison of average localization error for various algorithms with respect to sensor density in large scale simulation. $m = 20$**.**



(a) ITP fusion algorithm. $d = 7.5$ feet. (b) MoE (c) MLE

Figure 12: Average localization error with respect to different source strength. Sensor density = 0.01 sensor/m^2**,** $d = 7.5$ **feet,** $m = 20$**.**

We show that ITP improves upon MoE by (1) explicitly eliminating the effects of phantom estimates, and (2) preferring the higher SNR estimates among the non-phantom estimates. Lastly, we show that ITP is somewhat less accurate than MLE when n is large, but it can run significantly faster than MLE, whose performance is highly sensitive to initial estimates of the localization parameters.

8. ACKNOWLEDGMENTS

We wish to thank Frank DeNap, director of SensorNet project at Oak Ridge National Laboratory, for fruitful discussions and advice during our research efforts. Thanks are also due to the anonymous reviewers whose constructive comments have helped us improve the paper.

9. REFERENCES

[1] http://www.rftrax.com/radczt.html.

[2] X. Cheng, A. Thaeler, G. Xue, and D. Chen. Tps: a time-based positioning scheme for outdoor wireless sensor networks. In *INFOCOM 2004: Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 4, pages 2685–2696, 7-11 March 2004.

[3] J. Cox and M. B. Partensky. Spatial Localization Problem and the Circle of Apollonius. *ArXiv Physics e-prints*, Jan. 2007.

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, 1996.

[5] S. Guha, R. Rastogi, and K. Shim. CURE: an efficient clustering algorithm for large databases. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 73–84, 1998.

[6] A. Gunatilaka, B. Ristic, and R. Gailis. On localisation of a radiological point source. *IDC '07: Information, Decision and Control*, 2007.

[7] J. A. Hartigan. *Clustering algorithms*. Wiley, 1975.

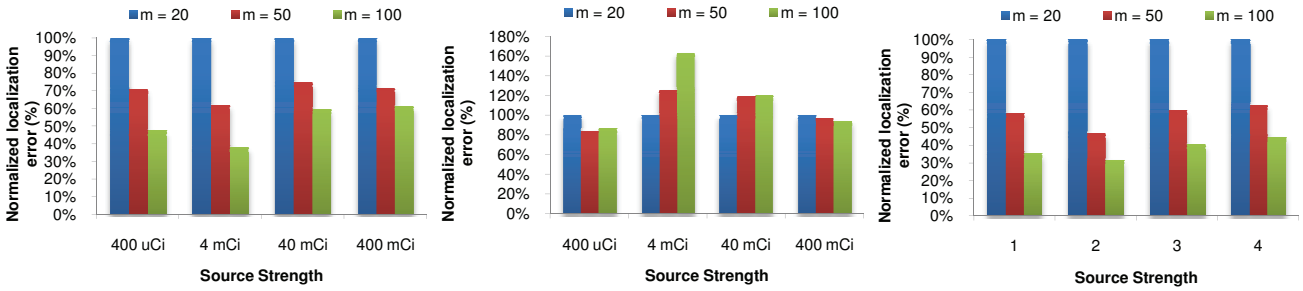
[8] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. Range-free localization schemes for large scale sensor networks. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 81–95, New York, NY, USA, 2003. ACM.

[9] L. J. Heyer, S. Kruglyak, and S. Yooseph. Exploring expression data: Identification and analysis of coexpressed genes. *Genome Research*, 9:1106–1115, 1999.

[10] G. F. Knoll. *Radiation Detection and Measurement*. John Wiley, 2000.

[11] D. Mihalas and B. W. Mihalas. *Foundations of Radiation Hydrodynamics*. Courier Dover Publications, 2000.

[12] D. Niculescu and B. Nath. Ad hoc positioning system (aps) using aoa. In *INFOCOM 2003: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies*, pages 1734–1743 vol.3, 2003.



(a) ITP fusion algorithm. $d = 7.5$ feet. (b) MoE (c) MLE

Figure 13: Normalized 90% percentile localization error for different values of m with respect to source strength. The localization error is normalized to the localization error of $m = 20$ and respective source strength. Sensor density = 0.01 sensor/m^2, $d = 7.5$ feet.

[13] C. S. Ogilvy. *Excursions in Geometry*. Dover Publications, 1990.

[14] N. S. V. Rao. Identification of simple product-form plumes using networks of sensors with random errors. In *International Conference on Information Fusion*, 2006.

[15] N. S. V. Rao, M. Shankar, J.-C. Chin, D. K. Y. Yau, S. Srivathsan, S. S. Iyengar, Y. Yang, and J. C. Hou. Identification of low-level point radiation sources using a sensor network. In *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2008.

[16] N. S. V. Rao, M. Shankar, J.-C. Chin, D. K. Y. Yau, Y. Yang, J. C. Hou, X. Xu, and S. Sahni. Localization under random measurements with application to radiation sources. In *International Conference on Information Fusion*, 2008.

[17] N. S. V. Rao, X. Xu, and S. Sahni. A computational geometry method for dtoa triangulation. In *International Conference on Information Fusion*, 2007.

[18] A. Savvides, C.-C. Han, and M. B. Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *MobiCom '01: Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 166–179, New York, NY, USA, 2001. ACM.

[19] G. Simon, M. Maróti, Ákos Lédeczi, G. Balogh, B. Kusy, A. Nádas, G. Pap, J. Sallai, and K. Frampton. Sensor network-based countersniper system. In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 1–12, New York, NY, USA, 2004. ACM.

[20] A. Sundaresan, P. K. Varshney, and N. S. V. Rao. Distributed detection of a nuclear radioactive source using fusion of correlated decisions. In *International Conference on Information Fusion*, 2007.

[21] A. Thaeler, M. Ding, and X. Cheng. itps: an improved location discovery scheme for sensor networks with long-range beacons. *Journal of Parallel and Distributed Computing*, 65:98–106, 2005.

[22] X. Xu, N. S. V. Rao, and S. Sahni. A computational geometry method for localization using difference of distances. *IEEE Transactions on Sensor Networks*, 2008. to appear.