Localization In The Presence of Spatially Obfuscated Sensor Reports

Supriyo Chakraborty and Mani B. Srivastava
UCLA
{supriyo,mbs}@ee.ucla.edu

Chatschik Bisdikian and Raghu K. Ganti
IBM T.J. Watson Research Center
{bisdik,rganti}@us.ibm.com

Abstract—Detection and localization of events is an important activity of military coalition networks characterized by distributed sensing and information exchange. However, the variability in mutual trust between members results in the formulation of sharing policies which often require reports to be deliberately obfuscated while maintaining the plausibility of their content. In this paper, using localization reports as evidences in support of (or against) hypotheses about event locations, we develop the foundations of an evidential reasoning-based approach that uses subjective logic for information fusion and inferencing for localization in the presence of incomplete and conflicting knowledge. To do so, we exploit our recent extensions of subjective logic that accommodate the spatial relationships that naturally exists between location reports from the different members. After highlighting our spatial extensions, we apply them in building an inferencing algorithm for a specific example scenario of primary user localization in a cognitive radio network. The reports are provided by various secondary users in the network. Through extensive simulations, we analyze the performance and the effect of various design parameters, showing a 90% accuracy in localization. Finally, we provide comparison results with other localization techniques via simulations.

I. INTRODUCTION

Detection and localization of events (e.g., sniper activity, armored vehicles) is an important activity for many military operations. It relies on distributed sensing and rapid dissemination of information to the operation participants. Aside from the inherent challenges in disseminating information over unreliable military networks, coalition settings, where the variability in mutual trust, in conjunction with heterogeneity in data handling and sharing policies between participants may result in information obfuscation, makes the dissemination of sufficiently useful information even more challenging.

We define obfuscation as a process which leads to degradation in information quality. It results from deliberate transformations applied with the aim of information hiding (at least to a certain degree) while maintaining sufficient information utility [4]. For example, for an event associated with parameters $<\text{type}, \text{location coordinates}>$, a sharing policy may dictate to accurately reveal the type of the event but selectively hide the location coordinates by providing only an obfuscated region. On one hand the multiplicity of sources provides sensing diversity, while on the other the increased uncertainty and possibly conflicting reports from the sources adversely affect the decision accuracy. Thus, we need techniques to explicitly quantify uncertainty and reason about the obfuscated data.

Specifically, our solutions approach is based on evidential reasoning about the location reports, i.e., the evidences, of an event location from sources of different trust levels that may also obfuscate their reports. In our case, this reasoning takes place against a backdrop of location hypotheses for the event and the outcome is an inference about the event location of high belief. Evidential reasoning is anchored on the Dempster-Shafer (DS) theory of evidence [14] which is a logic-based technique that deals with uncertainties in knowledge and allows reasoning in the presence of incomplete and conflicting pieces of evidence, as the case may be when obfuscation occurs. Evidential reasoning and DS theory has often been used for performing analysis (i.e., reason) of events observed during network operation for reasons such as anomaly detection [3] and network fault detection [17] or in intrusion detection [7] and DDoS [16].

In this paper we investigate the use of evidential reasoning in coping with the possibility of obfuscated location reports from sources in coalition networks. In doing so, we first adopt a generalization of the DS theory, that of subjective logic (SL) [9][10] that explicitly considers uncertainty that arises from the available evidences in addition to the belief and plausibility in DS. It provides a framework for subjective quantification of the belief, disbelief and uncertainty associated with a report when objective probabilities are hard to compute (due to inadequate information) [6]. However SL (and evidential reasoning) only deals with concrete (i.e., categorical) hypotheses, such as the state of a node being “good” or “bad” [17], a fault being of type “A”, “B” or “C” [1] etc. SL does not provision for the case where there are the spatial and, hence, measurable relationships between the hypotheses – a case which naturally arises when dealing with location reports and localization inferences. In response to this limitation, we have recently developed saSL, a spatial-aware extension to SL to allow consideration of “distances” between location hypotheses, comprising of spatial regions, and “influences” of one location reports to such hypotheses. In this paper, we highlight saSL and use it in developing an efficient window-based search algorithm, which finds the region of maximum belief as the event region.

To evaluate our algorithm, we consider an example of a cognitive radio network (CRN) which utilize radios that can change their transmission and reception parameters autonomically. This allows for improved utilization of licensed spectrum by allowing unlicensed users (the “secondary”) to access the licensed spectrum while avoiding potential conflicts with the licensed (the “primary” users) and other secondary users. Localization will typically involve location reports from secondary users about the location of a primary user. In
our evaluation, we study the localization accuracy vs. the trust that the location reports from the secondary users are accurate enough, while being subjected to varying levels of obfuscation. We show through simulation and experimental results that our algorithm can localize PUs with over 90% accuracy in the presence of obfuscated reports and performs better than traditional techniques such as Weighted Centroid Localization [20] applied to the same data.

In summary, the main contributions in this paper are: (a) the application of evidential reasoning and in particular of saSL, our spatial extension of subjective logic, to an event localization problem in coalition networks with sources of different trust levels that probabilistically obfuscate their location reports; (b) the introduction of a window-based search algorithm of complexity $O(\frac{\alpha^2}{\delta^2} + \frac{\delta^2}{\delta} + \frac{\alpha^2}{\delta})$ (where $\alpha$ is the dimension of the search area and $\delta$ is the step-size) that uses saSL to find the region of given size ($w$) that contains the event location with highest belief; and (c) an extensive simulation study of the algorithm performance using CRN as an example scenario and comparison with other localization techniques.

In Section II we describe our problem setting followed by Section III where we present a background on subjective logic and present out proposed spatial extension to it. The problem of activity region localization is described in Section IV followed by evaluation in Section V. Related work is summarized in Section VI. We conclude in Section VII.

II. SYSTEM MODEL AND PROBLEM DESCRIPTION

For ease of presentation, we consider an $a \times a$ square region $\mathcal{R}$ where the event occurs and the different sources reside. The region comprises of a grid of cells $(i, j)$, $1 \leq i, j \leq a$, see Fig. 1, and this grid is common to all the sources. An event $E := (ex, ey)$ is associated with location coordinates $1 \leq ex, ey \leq a$ in $\mathcal{R}$. It is detected by a subset of sources $S_E = \{s_1, s_2, \cdots, s_n\}$. The subset $S_E$ changes depending on where $E$ occurs as each source has a restricted sensing radius. Depending on the signal measured (e.g., RSSI, direction) each $s_i \in S_E$ makes an inference about the location of $E$ and generates a report $(R_i, t)$. The report $R_i$ is a *contiguous region* where $E$ could have possibly occurred and $t$ is the time corresponding to the measurement. In addition, each region $R_i = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$ is a set of $m$ cells such that $R_i$ is adequately obfuscated; in Fig. 1, $m = 4$ for each $R_i$.

For an event $E$, we associate a hypothesis $h_{ij}$ to each cell $(i, j)$ in the grid that the event has occurred in that cell; let $H$ be the collection of these hypotheses for all $i, j$, in $\{1, \cdots, a\}$. This effectively transforms the grid space into a hypothesis space within which a report can now be interpreted as an *evidence* in support of a subset of hypotheses. In other words, for every sub-region in $\mathcal{R}$ we can define a corresponding subset in $H$. Within the above framework, the effect of obfuscation can be modeled as an increase in evidential uncertainty and the problem of report fusion can be translated into one involving combination of uncertain evidences.

We consider a two-stage decision reporting scheme, where individual decisions from $n$ independent nodes are aggregated by a special node $D$, the decision maker, which performs fusion (see Fig. 2) of the reports and provides the final decision. As the figure also implies, along with each report, there is an associated level of trust for the source of that report. Similar to computational trust research, we interpret trust as a subjective probability with which a truster assesses that the trustee will perform a typical action. Essentially, trustworthiness evaluates both a trustee’s reliability as well as its regularity [5]. In this paper, we do not assume any specific trust assessment mechanism for a *trust block*. We simply assume that it associates a trust score $\theta(s_i, D) \in [0, 1]$, for $i = 1, \cdots, n$ for the trust the $D$ has on each source. We also assume a reciprocity in that the sources also maintain a trust on $D$ with a corresponding score $\theta(D, s_i) \in [0, 1]$, for $i = 1, \cdots, n$. In our evaluations, we assume that $\theta(D, s_i) = \theta(s_i, D)$.

Finally, the decision-maker $D$ receives a set of location reports $(R_1, t), (R_2, t), \cdots, (R_n, t)$ from sources of varying trust. Some of the reports could also be mutually conflicting, i.e., do not have regions of overlap. Thus, the problem at hand can be stated as: *Given a user-defined parameter $w$ and an event $E$ in location $(ex, ey)$, our objective is determine a contiguous region $R_f \subset \mathcal{R}$ of size $|R_f| = w$ that most likely contains the location of the event*. We define decision region $R_f$ to be successful when $(ex, ey) \in R_f$.

III. SPATIALLY-AWARE SL (SASL)

A. Background

SL is a framework for representing and reasoning about belief and uncertainty of hypotheses in the presence of evidences [9]. It was proposed as an extension to DS theory [14]. SL uses a *frame of discernment*($\Theta$) which is the set of all possible atomic hypothesis for a proposition. A source $s_i$, based on its observation can associate a belief mass $m_i(X)$ to any subset of $\Theta$ such that: (a) $m_i(\emptyset) \geq 0$; (b) $m_i(\emptyset) = 0$; and (c) $\sum_{X \in \Theta} m_i(X) = 1$. Belief mass is always associated is always associated to the entire set $X$ and not to any subset of $X$. The belief mass then used to compute the belief $b_i(X)$, disbelief $d_i(X)$, and uncertainty $u_i(X)$ functions over any region $X$. Collectively, the tuple $(b_i(X), d_i(X), u_i(X))$ is called the *opinion* of $s_i$ about set $X$ and $b_i(X) + d_i(X) + u_i(X) = 1$.

B. Case For Spatial Extension

Consider reports $R_1$ and $R_2$ in Fig. 1. Let $X \in \mathcal{R}$ be any region such that $R_1 \not\subseteq X$ and $R_2 \not\subseteq X$ then using the
the belief function of subsets. In addition, for any two regions \( X \) and \( Y \) consider non-empty subsets of the metric space. Each subset influences due to either reports at region \( Y \).

Thus, \((H, d_E)\) forms a metric space. However, in saSL, we consider non-empty subsets of the metric space. Each subset represents a collection of points in this space, and hence \( d_E \), which is defined between any two points, is not directly applicable. In addition, for any two regions \( X \) and \( Y \), a separation measure should also satisfy the following properties of the saSL framework.

\( P_1 \) If \( Y \subseteq X \), for consistency with SL, the entire belief mass \( m_i(Y) \) should be added to the belief function of \( X \).

\( P_2 \) If \( Y \supseteq X \), then only a fraction of \( m_i(Y) \) should influence the belief function of \( X \).

\( P_3 \) For a given region \( X \), the belief mass \( m_i(X) \) is always for the entire region and cannot be divided into any of its subsets.

\( P_1 \) and \( P_2 \) imply that we cannot use a symmetric metric. This rules out the possibility of using traditional separation metrics between subsets such as the Hausdorff distance [8]. For \( P_3 \), consider the following scenario where we want to compute the belief in \( X \) due to set \( Y \) with mass \( m_i(Y) \). If we take the maximum distance from any point in \( Y \) to \( X \) it results in the minimum influence of the mass \( m_i(Y) \) on \( X \). Correspondingly, for minimum distance we get the maximum influence. Also, the minimum distance is different for every point in \( Y \). Since we do not allow distribution of belief masses over any subregion of \( X \), we use an aggregate measure over all the point influences. Therefore, we take the above into account, and define the separation metric \( d_S(X, Y) \) for \( X, Y \subseteq H \) as:

\[
d_S(X, Y) = \frac{1}{|y|} \sum_{y \in Y} \min_{x \in X} d_E(x, y) \tag{2}
\]

The distance \( d_S(X, Y) \) is the minimum travel distance to reach \( X \) from \( Y \) averaged across all points in \( Y \); recall that a point represents a grid cell. Also, using Eqn. \( (2) \) we can show that \( d_S(X, Y) \geq 0 \) and \( d_S(X, Y) \neq d_S(Y, X) \), making \( d_S(X, Y) \) an asymmetric separation measure, reflecting the asymmetric relationship between \( X \) and \( Y \).

**D. Proposed saSL Definition**

We define an influence function \( I(d_S(X, Y); \beta) \) such that \( d_S(X, Y) \geq 0 \) and \( \beta \in \mathbb{R}_+ \) with the following properties:

\[
\begin{align*}
(I_1) & \quad d_S(X_1, Y) > d_S(X_2, Y) \Rightarrow I(d_S(X_1, Y); \beta) < I(d_S(X_2, Y); \beta) \\
(I_2) & \quad I(0; \beta) = 1 \\
(I_3) & \quad I(\infty; \beta) = 0
\end{align*}
\]

Property \((I_1)\) means that the influence of a region \( Y \) on another region \( X \) decreases monotonically with separation. Properties \((I_2) \) and \((I_3)\) state that the highest and the lowest influence occurs at the least and the greatest separation respectively. The parameter \( \beta \) controls the rate at which the influence decays. Candidate influence functions include \( e^{-\alpha d_S(X, Y)} \) and \( \frac{1}{d_S(X, Y)^2} \) for \( n \geq 1 \). We now define the elements of saSL.

**Frame of Discernment \((H)\)**: Similar to SL, the hypothesis set \( H \) is the frame of discernment.

**Belief Mass Assignment \( m_i(X) \)**: Interpretation is similar to SL. In our work, we use the trust score, which serves as a good indicator of past performance, to initialize the belief mass of a report. If \( Y \) is the report from source \( s_i \) with trust score \( \theta(s_i, D) \), then \( m_i(Y) = \theta(s_i, D) \) and \( m_i(H) = 1 - \theta(s_i, D) \). We assume that no node is fully trustworthy, hence \( \theta(s_i, D) < 1 \) and \( m_i(H) > 0 \).

**Belief Function \( b_i(X) \)**: To compute the belief function at region \( X \) due to the belief masses from a source \( s_i \), we sum the influences due to every non-zero belief mass at \( X \). Thus,

\[
b_i(X) = \sum_{Y \in 2^H, m_i(Y) \neq 0} m_i(Y) I(d_S(X, Y); \beta) \tag{3}
\]

**Disbelief Function \( d_i'(X) \)**: Since every mass influences the belief at \( X \), their contribution to disbelief is zero. Thus, \( d_i'(X) = 0 \).

**Uncertainty Function \( u_i'(X) \)**: It is defined as

\[
u_i'(X) = \sum_{Y \in 2^H, m_i(Y) \neq 0} m_i(Y) (1 - I(d_S(X, Y); \beta)) \tag{4}
\]

The opinion about set \( X \) due to a report from \( s_i \) is given by the tuple \((b_i(X), u_i(X))\) and as in SL \( b_i(X) + u_i(X) = 1 \).

**Fusion Rule**: To combine the opinion from two different sources \( s_i \) and \( s_j \) we use SL’s consensus operator [10] defined as follows:

\[
b_{i,j}'(X) = \frac{b_i(X) u_j'(x) + b_j'(X) u_i'(x)}{u_i(X) + u_j(X) - u_i(X) u_j(X)} \tag{5}
\]

\[
u_{i,j}'(X) = \frac{u_i(X) u_j'(x) - u_i'(X) u_j(X)}{u_i(X) + u_j(X) - u_i(X) u_j(X)} \tag{6}
\]

In the next section we apply saSL definitions to perform event localization.

**IV. Event Localization**

saSL allows us to compute the combined belief function, due to the reports, over any subset in \( H \). Our goal as part of the localization process is to find subset \( R_f \) in \( H \) of size \( 1 \leq w \leq m \) such that \( b_{1:2: \ldots : n}(R_f) \geq b_{1:2: \ldots : n}(R) \) for all \( R \subseteq H \) and \( |R| = w \). The set \( R_f \) corresponding to maximum
combined belief is reported as the event location. If we do not include partial cells in our selected subsets, then for $\mathcal{R}$, there are $\binom{n^2}{w}$ ways of choosing set $R_f$. Even if we constrain the selected region to be contiguous there remains a large number of possibilities. In this section, we exploit properties of saSL to reduce the size of the search space over which we estimate $R_f$ and then propose a search algorithm for finding it.

We define the notion of a bounding box $B$, illustrated in Fig. 1, as the smallest rectangular area covering all the reported regions. We use Lemma 1 and Lemma 2 to show that for any subset $X_o \notin B$ of size $w$, where $w \ll |B|$, we can find a subset $X_i \in B$ again of the same size such that the separation $d_S(X_i, R)$ is always less than or equal to $d_S(X_o, R)$ for any report $R$ in $B$. Hence, a subset $X_i$ completely inside $B$ is closer to the reports than any $X_o$ partially or completely outside $B$. We further use these Lemmas and saSL's combination rules to show that the combined belief in $X_i$ is greater than or equal to the combined belief in $X_o$, see Proposition 1. This, allows us to restrict the search space to only the bounding box $B$. However, if the sources are malicious $B$ could be as large as the monitored region $\mathcal{R}$. We finally propose a window-based heuristic search algorithm to scan $B$ and find set $R_f$ corresponding to the highest belief.

**Lemma 1.** Let $B$ be a bounding box in a plane and $p$ be a point not in $B$. Then there exists a point $q$ in $B$ such that for all points $r$ in $B$, we have $d_E(q, r)$ is at most $d_E(p, r)$.

**Proof:** See technical report for details [2].

**Lemma 2.** Let $B$ be a bounding box in a plane. We define $w$ such that $1 \leq w \leq m$ and $w \ll |B|$ where $m$ is the size of a report in $B$. For any set $X_o$ of size $w$ not completely in $B$ there exists an equal sized set $X_i$ completely inside $B$ such that for all reports $R$ in $B$ we have $d_S(X_i, R) \leq d_S(X_o, R)$.

**Proof:** See technical report for details [2].

**Proposition 1.** Let $B$ be the bounding box of the reports from $n$ different sources where each report is of size $m$. For any set $X_o$ not completely in $B$ of size $1 \leq w \leq m$ we can find an equal sized set $X_i$ in $B$ such that $b'_1:2:...:n(X_i) \geq b'_1:2:...:n(X_o)$ where the total belief is due to the reports in $B$.

**Proof:** We use Lemma 1 and Lemma 2 to show that for any set not completely in $B$ we can find a set inside $B$ which is closer to the reports. We compute the opinion due to each report on both the regions and then combine them to prove that the total belief at the region inside $B$ is greater than or equal to that at the region not completely in $B$. The details of the proof are in the technical report [2].

### A. Linear Search

We use a window of size $w$ cells ($\sqrt{w} \times \sqrt{w}$) to exhaustively scan the bounding box. We perform a horizontal scan by sliding the window in steps of $\delta$. When the window reaches the edge of the bounding box, it is shifted vertically by $\delta$ and the horizontal scan is repeated. For a square box of dimension $a \times a$, linear scan has a high time complexity of $O(\frac{n}{m})$.

### B. Hybrid Search

To reduce the time required to search using linear scan we employ a hybrid strategy. We do a greedy search over the initial region to quickly reduce the size of the search area and then apply linear scan over the smaller region. Algorithm 1 outlines the hybrid search. The step size $\delta$ is initialized to 1 and the number of cells to be searched $\text{numCells}$ is set to the total area on line 1. In each iteration, lines 3-4 initializes a vertical window $(\text{vwin})$. In lines 6-12 we perform a scan of the region in by sliding the window in steps of $\delta$ to compute the highest belief region as $\text{vwinMax}$. The minimum and maximum column boundaries $\text{cMin}$ and $\text{cMax}$ respectively are set on line 13. Lines 14-15 compute the robustness of the fusion scheme in the presence of malicious sources. Let an event occur in region $R_1$. There are $n$ sources which report the event of which $l$ are malicious. We index the malicious sources using $i = 1, \ldots, l$ and the trusted sources by $j = l+1, \ldots, n$. We define a source to be malicious if it colludes with other sources and reports a region $R_2$ different from the actual region $R_1$ such that $d_S(R_1, R_2) > d_m$ and $d_S(R_2, R_3) > d_m$ where $I(d_m; \beta) \approx 0$. We assume that the sources reporting region $R_1$ are equally and highly trusted and every report is assigned a belief mass of $m_j(R_1) = m_1$. Similarly, the reports from malicious sources are each assigned a belief mass of $m_i(R_2) = m_2$ where $m_1 \gg m_2$.

**Proposition 2.** The decision region would shift to region $R_2$, different from the actual region $R_1$, if and only if

$$l \geq \frac{n(m_1 - m_1 m_2)}{m_1 + m_2 - 2m_1 m_2}$$
where \( l \) is the number of malicious nodes reporting region \( R_2 \).

**Proof:** See technical report for the proof [2].

If \( m_1 = m_2 = m \), then Eqn. 7 reduces to \( l \geq \frac{2}{3} \). An intuitive interpretation is that when the mass clusters at regions \( R_1 \) and \( R_2 \) are widely separated then the decision would be incorrect only if greater than half the sources are malicious.

V. EVALUATION

For evaluation, we consider the localization of the primary user using reports from secondary users in a cognitive radio network (CRN). To increase the operational efficiency of such a network, the radios in a CRN need to determine under-utilized spectrum during periods of inactivity of the primary user (PU), thus an important operational aspect in CRNs is activity detection and localization of the PU [12][19] (event of interest). Typically, localization can be achieved by using reports generated by the network of secondary users (SU), corresponding to coalition members, that sense the channel for PU activity. SUs may obfuscate their location reports, i.e., deliberately alter the content of the reports by, for example, reducing the granularity of the location, translating the location, etc., [22], which could result in disruption of network operation including illicit acquisition and use of the licensed spectrum [11][18].

In the following, we describe our simulation setup and report results for the different experiments that we performed. We then provide comparison results for our technique with traditional localization techniques.

A. Setup

Our simulation setup consists of a square region \( \mathcal{R} \) of dimension \( a = 100 \) units divided into 10,000 cells of unit area. Events are generated at random locations in \( \mathcal{R} \). There are \( n = 20 \), SUs of varying trust scores reporting the event location. The reports are each obfuscated and contain \( m \) cells (square region of dimension \( \sqrt{m} \) cells). As stated in Section II, a decision is successful if the event coordinates \((ex, ey)\) lie in the decision region \( \mathcal{R}_f \). We define percentage of successful decisions as accuracy and use it to quantify our results. A summary of the different parameters is given in Table I. We report variation in accuracy for various combinations of these parameters. In all our experiments we set \( \beta = 1 \) so that the influence decays only as a function of separation.

**Trust Scores:** For generating the distribution of trust scores for the SUs we use a combination of parameters \( trustTh \) and \( \gamma \). For a given value of the trust threshold \( trustTh \), we generate trust scores in \([0, 1]\) such that \( \gamma n \) SUs have values higher than \( trustTh \) (trusted nodes) and \( n - \gamma n \) have lower values than \( trustTh \) (untrusted nodes).

**Obfuscation Model:** We build on the obfuscation model given in [15]. We use a composition of scaling and translation for location obfuscation. The obfuscation process takes as input three parameters: scaling \( m \), translation \( \alpha \) and trust score \( \theta(D, s_i) \). While every report is scaled to include \( m \) cells, translation depends on the trust score. We define a Bernoulli r.v. \( X \) with \( p = 1 - \theta(D, s_i) \) and use it to determine the reports to be translated. The translation to coordinates \((ex', ey')\) is given by:

\[
\begin{align*}
ex' &= ex + XZ_1 \quad (8) \\
 ey' &= ey + XZ_2. \quad (9)
\end{align*}
\]

Values for \( Z_1 \) and \( Z_2 \) are drawn from a Gaussian distribution \( N(\mu = 0, \sigma^2 = \alpha(1 - \theta(D, s_i))) \). The translated point \((ex', ey')\) is then scaled. During scaling we ensure the following: First, the chosen cells are contiguous. Second, if a report contains the event then the containing cell is distributed uniformly within it. Note, while \( \alpha \) is constant for every SU the parameter \((1 - \theta(D, s_i))\) makes the obfuscation SU specific.

B. Report Characterization

**Mean Separation:** We use separation measure \( d_S(.) \) to quantify the effect of both scaling and translation on a report. Thus, even when there is no translation (i.e., \( X = 0 \), we have a non-zero separation due to scaling. We generate 1000 random events and corresponding reports with a fixed value of \( m \). The mean separation \( \langle d_S(E, Y) \rangle \) between the event location \( E \) and generated report \( Y \) for varying values of \( \alpha \) and \( \theta(D, s_i) \)
and report size value of $K$ values of $K$ and plot the average size of the bounding box for varying completion time. We fix $\gamma K \leq \lceil m \right)$ expected as an increase in $C$. Accuracy included.

reports are clustered and is observed when trusted sources are obfuscated. The increase in the box size is pronounced for the bounding box increasing the localization uncertainty. From our observations, setting $\sqrt{\gamma w}$ to 225 provides a good tradeoff between accuracy and uncertainty in decision region.

Variation with $\alpha$ and $\gamma$: Together the value of $\theta(D, s_i)$ determines the distribution of the trust scores. For a given $\theta(D, s_i)$, we vary the value of $\gamma$ and illustrate the accuracy in Fig. 6. For a fixed $K$ and $\theta(D, s_i)$, as we increase $\gamma$ more trusted nodes are selected for decision making. The reports are typically clustered around the event, increasing the accuracy of localization. Further, for a given value of $\gamma$ an increase in the value of $\theta(D, s_i)$ increases accuracy. This is because the selected nodes obfuscate with a lesser probability and also with a lower magnitude. This helps in achieving a higher accuracy. An interesting observation is that for a high trust threshold $\theta(D, s_i) = 0.9$ and $\gamma = 0.3$, the trusted nodes selected are $\frac{K^2}{2} + 1$ and even then we can achieve an accuracy of around 80% which validates the robustness of the scheme as pointed in Proposition 2.

Variation with $K$ and $\alpha$: We set a high trust threshold $\theta(D, s_i) = 0.9$ and vary $K$ such that only trusted nodes are selected. In Fig. 7 we observe that for a given value of translation $\alpha$, the accuracy increases to higher than 90% with $K$. The influence function chosen decays inversely with separation. However, for a fixed $K$ an increase in $\alpha$ does not lead to a noticeable decrease in accuracy. We attribute this to two factors. One, the high value of $\theta(D, s_i)$ implies that the r.v. $X$ in the obfuscation model, is mostly zero and few nodes actually obfuscate. Second, the influence function reduces the effect of an obfuscated report thus minimizing its effect on the accuracy. Fig. 8 repeats the experiment for an exponentially decaying influence function. We observe that the fusion process performs well under both a fast decaying exponential influence and a relatively slow inverse decay function.

D. Comparison

For different obfuscation values we compare the accuracy of saSL against two different strategies.

Maximum Intersection Region: We employ a naive strategy. For each obfuscated report we add a weight equal to the trust score of the source to every cell contained in the report. This
is similar to a voting scheme, where each report votes for the cells it contains. The value of a vote is the trust score of the source providing the report. We choose the cell corresponding to the maximum value, breaking ties by random selection. We then report a region of size \( w \) containing this cell as our decision region.

**Weighted Centroid Localization (WCL):** It is a range-free technique which uses RSSI measurements for PU localization in cognitive networks. The advantage of a range-free technique is that error-prone modeling and estimation of channel parameters can be avoided. We apply the WCL algorithm [20] on the median of every report and weigh them according to their trust scores to obtain the location estimate. We then form a region of size \( w \) containing the estimated location and report it as the decision region.

The comparison results in terms of accuracy are shown in Fig. 9. saSL consistently outperforms both the strategies as it is able to handle uncertainty of a report and its influence on the neighborhood better.

VI. RELATED WORK

Detecting and isolating malicious secondary users and their data has seen a lot of interest in recent years. In [23] a reputation score is assigned to SUs depending on the level of agreement of their reports with others. SUs with reputation score below a threshold are ignored. In [21], an “onion-peeling” approach is followed to defend against multiple malicious nodes in a collaborative centralized sensing setup. Every SU is assigned a suspicious level, which is a possibility that the node is malicious, according to their reports. If an SU is found to have a suspicious level greater than a threshold it is tagged as malicious. Similarly, in [24] the authors propose a trust-based framework to secure the information aggregation process in the presence of malicious nodes. They exploit the multiplicity of nodes to extract underlying statistical characteristics from the sampled data and the use the “distance” between a measurement and these statistical characteristics to assign reputation and isolate malicious nodes. While quite motivated by these past works, our research is distinctly different from the past studies in that due to the probabilistic nature of the malicious behavior we do not explicitly try to isolate specific users. Instead we assess the influence of nodes and their evidences and appropriate weigh them to perform localization. Closest to our attack model is the work in [13]. However, the authors consider the problem of PU detection in the presence of malicious nodes. They use Bayesian reasoning and implement it using belief propagation on factor graphs.

VII. CONCLUDING REMARKS

We proposed a spatial separation based extension of subjective logic as a fusion mechanism for event localization using obfuscated and possibly inconsistent information collected from various sources in a coalition network.

Specifically, as an example scenario, we investigated the case of performing PU location estimation in the presence of secondary users that obfuscate (scale and translate) location reports. We have used the location reports from different users, each with a different trust level, as evidences in support of hypotheses about the PU location. Exploiting the saSL extensions of argumentation with subjective logic that we developed that account for the spatial relationship between evidences and the various location hypotheses, we have developed an evidence fusion and inference mechanism that finds the region of specified size that contains the PU with maximum belief. Extensive simulations show the robust behavior of the algorithm even in the presence of misbehaving sources and its improved performance when compared with other localization techniques for cognitive radio networks.

Having established and analyzed the fundamental principles of the localization mechanism based on evidential reasoning and saSL in this paper, future research plans include investigating the dynamic operation of the system where, for example, trust score for various sources evolve from one localization operation to the next; adapting the saSL parameters, such as those defined for the influence functions, to the trust scores and how these are updated; as well performing further evaluating the behavior of our algorithms against real data localization traces.

ACKNOWLEDGEMENT

The authors are thankful to George Tychogiorgos for help with the simulation study.

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defense under Agreement Number W911NF-06-3-0001 and by the NSF under award
REFERENCES


