IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 3, 2021, accepted April 13, 2021, date of publication April 20, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074249

A Crowd-Based Efficient Fault-Proof Localization System for IoT and MCS

ADARSH GHIMIRE, SELINA SHRESTHA, AND HADI OTROK, (Senior Member, IEEE)
Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Corresponding author: Hadi Otrok (hadi.otrok@ku.ac.ae)

This work was supported in part by Khalifa University, and in part by the Abu Dhabi Department of Education and Knowledge (ADEK) under Grant AARE18-106.

ABSTRACT In this paper, an efficient and fault-proof active node selection approach for localization tasks in Internet of Things (IoT) and Mobile Crowd Sensing (MCS) systems is proposed. The proposed approach is resilient to the presence of anomalous nodes. Localization is the process of fusing data readings from multiple sensing nodes with the aim of finding the location of a specific target, such as radiation source, forest fires and noisy areas. Current active node selection systems for localization tasks perform algorithms like greedy and genetic methods over the whole Area of Interest (AoI). As such, a system that considers anomalous data is required to detect anomalies and perform localization over a large number of active nodes, which usually takes multiple iterations and is computationally costly. To overcome this, we propose a resilient localization approach which a) uses the median filter based image filtering technique to level out anomalous readings, b) uses the filtered readings to reduce the AoI to be around the target location without being influenced by anomalous nodes, c) detects and eliminates anomalies in the new AoI based on the deviation between filtered readings and original readings, and d) selects remaining nodes in new AoI for localization. As a result, there is a huge reduction in the complexity of active node selection and thus reduction in time taken by the system to perform the task of source localization. The efficacy of the proposed system is evaluated for radiation source localization tasks using simulated radiation dataset, by performing experiments for several test scenarios. The results demonstrate that the system is able to perform localization tasks in significantly reduced time and therefore generate near real-time results while also maintaining low localization error.

INDEX TERMS Anomaly detection, image filtering, Internet of Things, localization, mobile crowd sensing, weighted average.

I. INTRODUCTION

The Internet of Things (IoT) allows the connection and interaction among different objects like sensors, RFIDs, actuators, and mobile phones, due to which it is widely being considered as the next big technological revolution [1], [2]. The paradigm of IoT enabled the concept of Mobile Crowd Sensing (MCS), in which a number of mobile devices act as sensing nodes and collectively share sensing data so as to measure or predict a phenomenon of common interest [3]–[5]. Owing to the ubiquity of smart devices and the development of information technologies, IoT and MCS have gained huge popularity in the field of data collection [6], [7], and are thus well known areas of research in the present day.

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Verticale.

Target and event localization is one of the many applications of IoT and MCS. Localization tasks integrate and analyze data readings from multiple sensing nodes to identify the location of a particular event or phenomenon within a certain AoI. These can be used to localize events or phenomena like radiation, forest fires, noise pollution, water pollution, air pollution, etc [8]–[11]. It has been shown that the localization performance is considerably enhanced by employing many well-distributed small sensors, rather than using a few big ones [8]. The sensing nodes, however, comprise of small battery powered devices with limited residual energy, resulting in a need to minimize the number of nodes used to perform localization effectively. As such, optimum localization and optimum node selection/ placement are both active areas of research in localization tasks. In addition, these networks may comprise faulty nodes, which if used may deteriorate the system's localization accuracy. Thus, there is also a need for a mechanism that ensures robustness over such anomalous nodes.

Current works fuse data from multiple sensors for reliable outcomes. Sensor optimization methods in the literature primarily consist of 2 types: node placement optimization [12]–[14] and active node selection [15]–[19]. The node placement optimization methods are concerned with placing the sensing nodes in a certain AoI with the objective of minimizing the localization time and the number of deployed nodes, as well as maximizing the localization accuracy. Likewise, the active node selection techniques focus on selecting the best group of sensing nodes from the available population by considering several selection parameters such as cost, area coverage, node's residual energy, efficiency, and reputation. The former is used in relatively earlier works and are not as effective since the prior knowledge of the source location is unknown, making the method complex and less adaptive.

Localization tasks rely heavily on sensor readings to estimate the target location. However, sensing nodes may sometimes provide incorrect readings due to malfunctioning of its components or maliciousness. Such anomalous readings, if used, introduce very high errors in the target estimation, which can be extremely catastrophic in sensitive applications. Thus, detection and removal of such node readings is very important for accurate processing of such information. Most of the existing localization systems do not take anomalous readings into account and thus are less applicable to real life scenarios [8]–[12], [15], [20]–[25].

Most of the common anomaly detection systems are based on clustering, nearest neighbor, and, classification [26], [27] which detect outliers that differ from the majority. However, data points in localization tasks naturally vary from each other, even though they are in the same neighborhood. They follow a radial pattern depending on the node's distance from the event, where the nodes in closer proximity to source show higher readings. Therefore, these methods are not effective in localization tasks. A data-based centroiding technique for the anomaly detection for localization tasks was introduced in [16]. However in this work, the anomaly detection as well as localization is done over the whole AoI and therefore takes multiple iterations for completion. This is not ideal for time and resource constraining tasks. The major shortcoming of this method is that the localization process requires multiple rounds of node selection and localization, which: a) is computationally expensive, b) takes longer time to localize.

To overcome the aforementioned shortcomings, an Efficient Fault Proof Localization System (EFLS) has been proposed in this work. The system aims to reduce the complexity and time required to localize by narrowing the AoI to a small area around the target location. Before narrowing the AoI, correctness in the presence of anomalous nodes is ensured in a novel way by leveling out the anomalous data readings using noise filtering techniques in images [28]–[30]. The filtered readings are then used to reduce the AoI around high valued nodes. As opposed to complex and iterative methods that are performed over the whole AoI in existing works, EFLS presents a simpler and faster method to select active nodes in the presence of faulty readings. EFLS thus results in: a) reduced complexity, b) reduced time to localize, c) anomaly handling.

Thus, in comparison to current works, the major contributions of the proposed work are:

1) The *design* of a novel mechanism using median filter based image filtering technique to minimize AoI around actual target in the presence of anomalous nodes;
2) The *design* of a novel anomaly detection technique using deviation between actual and filtered data;
3) Reduction in complexity of active node selection by significant reduction in AoI;
4) Reduction in total time taken for localization and generation of near-real time results without compromising task accuracy.

The efficacy of the proposed system is evaluated by using real-life radiation datasets, where the task is to localize the radiation source. However, it can also be applied to other event localization tasks such as noise, fire, etc.

The rest of the paper is organized as follows. Section II goes over published works on related topics. Section III discusses the proposed approach consisting of median filtering, AoI reduction, anomaly detection and localization. Section IV presents and discusses the simulation results and evaluation metrics used. Finally, section V concludes the paper.

II. RELATED WORK

Many researches have taken place in literature which are aimed towards developing algorithms to solve the problem of localization using information from a network of sensing devices that is distributed in a particular area. Some works from literature [8], [12], [21]–[24] are mainly concerned with just computing the mathematical estimation of the target location based on the data from sensing nodes, while other works [12]–[16], [25] focus on development of full localization system which addresses the optimal node selection for the purpose of optimizing the task of localization.

Most of the early researches on source localization are primarily aimed at minimizing the localization time and error. [23] uses the Iterative Pruning Method to fuse data from sensing nodes for localization, while [22], [31] use Maximum Likelihood Estimate, and Inverse Square law to estimate the target in the neighborhood of the global maxima. [21], [23], [24], [32] use Time Difference of Arrival(TDoA) and Direction of Arrival(DoA) on readings to find the source location. Likewise, Bayesian method is used by [8], [12], [15], [16], [31], [33], in which prior belief about the target location are updated based on the fusion of the current data readings.

In addition to performing localization, some of the early works have also focused on finding the best location to deploy the sensing nodes. Since this is a NP-hard problem, most works have used greedy and genetic algorithms. [12] has proposed a solution using 9 stationary nodes while [13], [14] have considered placement of mobile nodes. These techniques targeted to minimize the number of nodes used along with minimizing the localization time and maximizing the localization accuracy. However, these node placement techniques require prior knowledge of the event location, and are thus not practical. More recent works have thus focused on selecting a number of nodes from the whole population to use for localization. A data-driven active node selection mechanism while considering parameters including residual energy, data confidence, area coverage, and cost is used in [15], [16], [19]. Likewise, [17], [18] perform an initial estimation of the target to generate the distance weight measurements for active node selection.

Most of the aforementioned works, however, do not incorporate a way to deal with situations when the sensor readings are anomalous, making these techniques highly prone to erroneous results in the presence of abnormal data. In literature, there are several anomaly detection techniques which are very accurate depending on the type of data used [26], [27]. Statistical modelling develops a statistical model to capture the distribution of data, and flags a data instance as anomalous if it does not fit the distribution [34]–[36]. Nearest neighbor technique flags anomalies based on their distance from the neighborhood [37]–[40]. Similarly, clustering technique flags anomalies if the data instance lies far from the cluster [41], [42]. Classification techniques involve a model that can classify if the instance is anomalous or normal by learning from training instances [43]–[45]. However, these anomaly detection algorithms give poor results for detecting anomalies in radially spreading data of localization tasks as they cannot capture the distribution of the radially spreading data. A data-based centroiding technique has been proposed in [16] to effectively detect anomalous readings in localization tasks. This work implemented an end-to-end localization system using the data driven active node selection based on greedy and genetic algorithms and Bayesian localization algorithm.

In this paper, we propose an efficient fault-proof localization system that handles anomalous readings effectively. The system uses fast image noise filtering techniques to outlie anomalous readings, and make the localization process more efficient. Previous works in the literature have never analysed anomalies in localization tasks from the perspective of image noise filtering. Noise that can exist in images include gaussian noise, impulse noise, shot noise, uniform noise, film grain noise and multiplicative noise. These noises can be refined perfectly by using linear filters or non-linear filters or combination of those filters to obtain noise free images [28]–[30]. In the case of localization tasks, node positions and node readings can be considered as analogous to pixel positions and pixel intensities respectively in images. As such, anomalous readings in sensing nodes can be inferred to be analogous to noise in images. The anomalies in localization tasks resemble the salt and pepper or impulse noise in images the most. These types of noise in images have been shown to be effectively removed by applying median filtering technique [28]–[30]. Similar technique, when applied in distributed node readings, gives a filtered data that is free of anomalous values, which is beneficial for efficient localization and anomaly detection.

TABLE 1. List of abbreviations.

abbreviation	Term
AoI	Area of Interest
CPM	Counts Per Minute
DoA	Difference of Arrival
$EFLS$	Efficient Fault Proof Localization System
IoT	Internet of Things
KNN	K-Nearest Neighbors
MCS	Mobile Crowd Sensing
MLE	Maximum Likelihood Estimate
$TDoA$	Time Difference of Arrival

III. PROPOSED SYSTEM

In source localization problems, the data readings provided by the sensing nodes deployed in the AoI are aggregated and processed with the aim of estimating the source location. This can be done efficiently and accurately by using sensor readings of nodes close to the actual source. In addition, faulty readings can inject high errors in target prediction, due to which selection of nodes closer to the source that are not faulty has to be done. In this work, radiation source localization is used as a running example. Given a single radioactive source, the goal is to use the sensor readings of nodes close to the actual target in order to predict the source location while considering anomalous readings.

With the goal of minimizing complexity of node selection process in a fault-proof manner, the proposed localization system makes use of an image filtering technique: the median filter, to regularize anomalous readings so that the Area of Interest (AoI) can be narrowed down without being influenced by anomalies. The system works in 5 main stages: 1) Data Collection and 2-D Mapping, 2) Filtering, 3) AoI Reduction, 4) Anomaly Detection and Elimination, 5) Localization. The sensor readings are first collected and mapped into a 2 dimensional grid of readings spaced according to node positions in the AoI. This 2-D image-like map is then passed through a filter to level out anomalies, the result of which is used to reduce the AoI. Nodes within the new AoI are then evaluated to detect anomalies. Anomalous nodes are eliminated and localization is performed over the remaining nodes in the new AoI. Figure 1 shows the system overview.

A. DATA COLLECTION AND 2-D MAPPING

The readings from deployed sensing nodes along with their locations and node IDs are obtained at once. This list of node coordinates and their sensor readings are then converted into a 2 dimensional grid. This is done because the image filtering stage only takes a 2 dimensional data that resembles an image. The position of a node in the 2-D grid is determined using its location coordinates and the value at that position is given by the node reading. As such, a node position in the grid

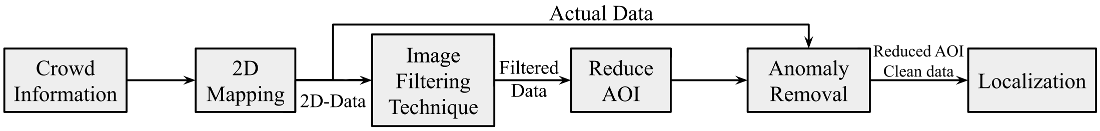
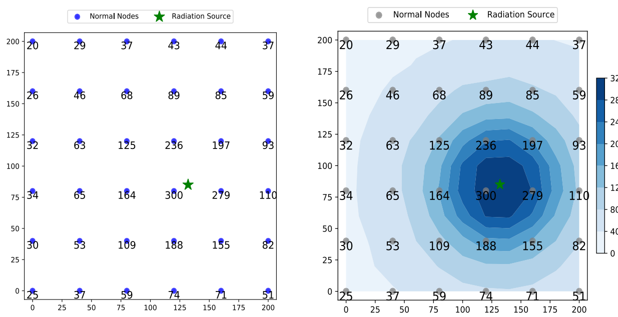


FIGURE 1. Proposed system block diagram.

TABLE 2. Example of collected data readings.

ID	X	Y	Sensor Reading
1	0	0	25
2	20	100	46
3	40	120	63
4	60	180	41
5	80	80	164
6	100	80	259
7	120	40	188
8	140	20	113
9	160	80	279
10	200	160	59
.	.	.	.
.	.	.	.
.	.	.	.



(a) Two dimensional map **(b)** Contour plot of sensor readings

FIGURE 2. Two dimensional representation of the data collected along with corresponding contour plot.

is analogous to a pixel in images and a node reading to a pixel intensity. Table 2 shows the collected data readings and Figure 2a shows the 2-D map for this data. Figure 2b is the contour plot of sensor readings in Figure 2a. Contour plot has been used throughout this paper in order to better visualize the magnitude of data values in the two dimensional space. In this scenario, the readings are collected through radiation detectors regarding a radioactive target that is present in the AoI as shown in Figures 2a and 2b. Details about the simulation environment are to be discussed in section IV-A.

B. FILTERING

In the domain of digital image processing, different filtering techniques exist for different types of noises. One of these noises is the impulse noise (also known as salt and pepper noise), which are sharp, random changes in some image pixel intensities. The anomalies in localization tasks resemble this noise the most. This type of noise in images has been shown to be effectively removed by applying the median filtering

(a) Noisy Coin Image **(b)** Median Filtered Coin Image

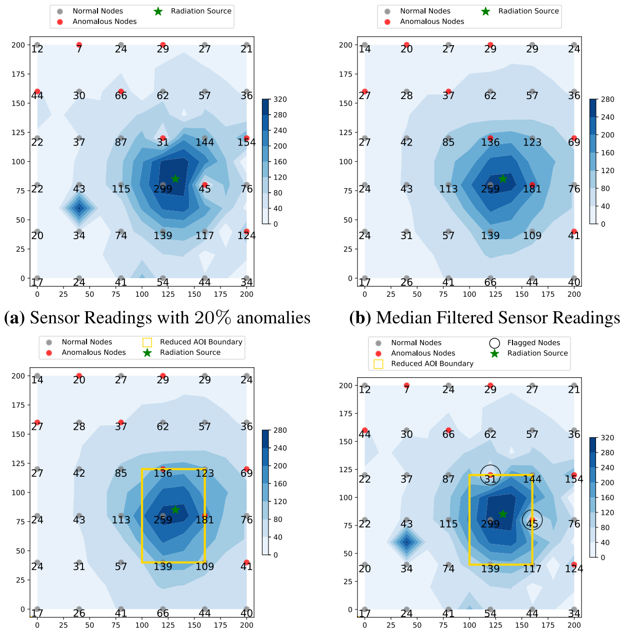
FIGURE 3. Impulse noise image before and after filtering [28].

technique [28], [29]. Figures 3a and 3b show an image with impulse noise and the same image after applying median filter respectively, where the noises have been removed effectively.

Therefore, the same technique has been used in the proposed system in order to filter anomalous readings. A median filter of size 3×3 has been glided through the grid of sensor readings with a stride length of 1. A 3×3 filter is suitable to filter out moderate density noise without blurring the output [46] and stride 1 ensures minimal information loss. For each filter operation, the filter outputs the median value of the 9 nodes under consideration. Since median values are not influenced by outliers, anomalous node readings are effectively replaced by the median value of the surrounding nodes in the filter output. This results in a filtered grid of data where the anomalous readings have been leveled out in accordance to their surroundings. This in turn results in a radially spread data which maintains the pattern and target position of the original readings. Figures 4a and 4b show node readings with anomalous values and the median filtered node values respectively. The anomalous readings have been leveled out effectively in the filtered data. For instance, the node at coordinate (120, 120) shows an anomalous reading of 31 in Figure 4a, while the surrounding nodes show much higher readings. This value has been leveled out by the median filter in Figure 4b to be 136, which is closer to the surrounding node readings.

C. AoI REDUCTION

The AoI is reduced with the objective of having a few number of informative nodes around the actual target owing to the fact that the use of active nodes closer to the target results in a faster and more accurate localization [8]. Since anomalous readings are nullified by the filter, the filtered data can be



(a) Sensor Readings with 20% anomalies (b) Median Filtered Sensor Readings

(c) Filtered Readings with generated boundary of the new AoI (d) Anomalous Nodes flagged within new AoI

FIGURE 4. Selection process of active nodes closer to the source.

used as a reference to interpret the rough target location in a way that is resilient to the presence of anomalous nodes. Thus, the AoI is reduced to be around the nodes with high filtered values. In the proposed system, this is done according to Algorithm 1.

Algorithm 1 Area of Interest Reduction

Result: Reduced AoI consisting of nodes closer to the target

Sort the filtered readings in descending order

Select top 10 % of filtered sensor readings

Compute boundary of new AoI as:

  Minimum and Maximum x coordinate of top 10% nodes

  Minimum and Maximum y coordinate of top 10% nodes

Select nodes inside the boundary of new AoI

Figure 4c shows the generated boundary of the reduced AoI for the filtered data in Figure 4d.

D. ANOMALY DETECTION AND ELIMINATION

Although several anomaly detection techniques like statistical modelling, nearest neighbor, clustering, and classification-based techniques exist in the literature, they do not work well with the radially spreading data of localization tasks. So, in this work, we propose a novel mechanism to detect anomalies within the reduced AoI by comparing the value of actual and filtered data for each node. As mentioned in section III-B, the anomalous data points undergo drastic value updates on filtering while the other data points undergo little changes as per their surroundings. Thus, we use this deviation to flag anomalous nodes and eliminate them according to Algorithm 2. Figure 4d shows that the anomalous nodes within the new AoI for sensor readings of Figure 4a are detected correctly.

Algorithm 2 Anomaly Detection and Elimination

Result: Anomaly flagged and removed inside the new AoI

For each node in the reduced AoI:

  Compute the deviation in actual sensor reading and filtered data

  $Deviation = ActualData - FilteredData$;

  Anomaly Flagging:

    If $|Deviation| \geq 0.5 \times ActualValue$:

      Flag node as Anomalous

    Else:

      Flag node as Normal

Remove Flagged Anomalous Nodes from AoI

E. LOCALIZATION

The efficient way to do the target localization is by selecting only those nodes that are closer to the source, and then executing the localization algorithm around the space of interest. The proposed system efficiently selects the nodes from the reduced AoI which are not anomalous as mentioned in section III-B and III-D, thus the next task is to implement the localization algorithm using the selected nodes only. The localization algorithm that has been devised consists of 4 main stages: 1) Hypothetical Space Generation, 2) Inverse Square Computations, 3) Similarity Computations, and 4) Weighted Average Computations.

1) HYPOTHETICAL SPACE GENERATION

In order to locate the target position, the reduced AoI is further divided into n_p uniform grids called patches. Central position of each patch is considered as a candidate location where the target might be situated. For localization, the possibility of the target being at each patch is accessed. Figure 5 shows the conversion of a reduced AoI with 5 sensing nodes ($n_s = 5$) into 9 hypothetical spaces where the target is possibly residing ($n_p = 9$).

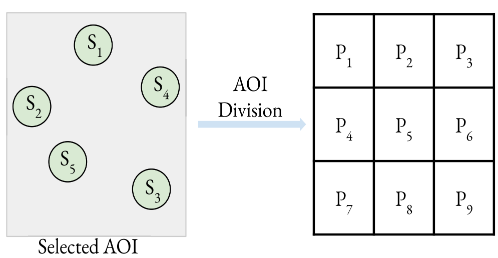


FIGURE 5. AoI to hypothetical patches.

2) INVERSE SQUARE COMPUTATIONS

Readings in localization tasks depend on the distance between the node and the target. In many phenomenons such as radiation, heat, and sound, the physical quantity at a certain location is inversely proportional to the square of distance between that location and the source of the physical quantity. Thus, in order to evaluate the likelihood of each patch being the source of the localization problem at hand, we assume each patch as the source and compute inverse square distance values from that patch to all the sensor nodes in the new AoI. The inverse square distance is computed using (1) and (2). The inverse square distances from a single patch to all the nodes in the new AoI are arranged in a one dimensional array which is referred as the distance vector.

$$ISD_{ij} = \frac{1}{d_{ij}^2} \tag{1}$$

where ISD_{ij} is the inverse square distance between location i and j and d_{ij} is the euclidean distance between location i and j computed using (2).

$$d_{ij} = \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2} \tag{2}$$

where (X_i, Y_i) and (X_j, Y_j) are the coordinates of locations i and j respectively.

The distance vectors of all the patches are arranged in a two dimensional array of size $n_p \times n_s$, such that each row represents the distance vector of a single patch. This two dimensional array is called the distance vector array and is computed using Algorithm 3.

Algorithm 3 Distance Vector Array Computation

Input: sensor positions (S_p), patch positions (P_p)
Output: distance vector array

```
Initialize DistanceVectorArray as an empty list
For i in each P_p:
    Initialize DistanceVector as an empty list
    For j in each S_p:
        Compute ISD_ij
        Append result to DistanceVector
    End For
    Append DistanceVector to DistanceVectorArray
End For
```

3) SIMILARITY COMPUTATIONS

To access which patch position is most likely to be the actual target, cosine similarity is computed between each patch distance vector and the vector of original sensor readings. Cosine similarity measures the similarity between two non-zero vectors, where 1 means they are maximally similar (i.e. parallel) and 0 means they are maximally dissimilar (i.e. perpendicular). The cosine similarity between two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is given as:

$$Cos(\boldsymbol{A}, \boldsymbol{B}) = \frac{\boldsymbol{A}.\boldsymbol{B}}{||\boldsymbol{A}||.||\boldsymbol{B}||} = \frac{\sum_{i=1}^{n} A_i.B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3}$$

where A_i and B_i are the i^{th} components of vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively and n is the length of vectors $\boldsymbol{A}$ and $\boldsymbol{B}$.

Thus, the cosine similarities between patch distance vectors and sensor readings imply the confidence score for each patch's proximity to the target. For example, if patches P_1 and P_2 have similarity scores of 0.9 and 0.85 respectively, then it implies that the target is 0.9 times closer to P_1 and 0.85 times closer to P_2. Similarly, if the similarity score of a patch is close to 0, then the possibility of the target being close to that patch is very low. Thus, to obtain the sweet spot between the likely patches, the weighted average algorithm is implemented.

4) WEIGHTED AVERAGE COMPUTATIONS

In order to find the location where the target is most likely located, the weighted average of the patch positions ($X_{wt.avg}$, $Y_{wt.avg}$) is computed using their similarity scores as:

$$X_{wt.avg} = \frac{\sum_{i=1}^{n_p}(X_{P_i} \times S_i)}{\sum_{i=1}^{n_p} S_i} \tag{4}$$

$$Y_{wt.avg} = \frac{\sum_{i=1}^{n_p}(Y_{P_i} \times S_i)}{\sum_{i=1}^{n_p} S_i} \tag{5}$$

where X_{P_i}, Y_{P_i} and S_i are the i^{th} patch x coordinate, y coordinate and similarity score respectively.

So as to avoid errors in target estimation introduced by including low score patches, the weighted average is computed only over the patches having a similarity score greater than 0.5. However, in cases with very high percentages of anomalous nodes, all patches may have a similarity score below 0.5. To account for this, if the maximum similarity score is 0.5 or less, then all patches are taken into consideration for weighted average computation. The computation of weighted average is shown in Algorithm 4 and the resulting location is the system's estimated target location.

Algorithm 4 Weighted Average For Target Location Estimation

Input: patch positions, similarity score
Output: Estimated Target Location

```
Find maximum similarity score
If maximum similarity score > 50%:
    Select patches with similarity score > 50%
Else:
    Select all the patches
Compute Weighted Average (X_wt.avg, Y_wt.avg)
```

IV. SIMULATION RESULTS

This section describes different experiments and simulations conducted to evaluate the performance of the proposed system: Efficient Fault-proof Localization System (EFLS).

A. SIMULATION ENVIRONMENT

The experiments are performed for a radiation source localization task. The radiation data readings are simulated using the principles of radiation, according to which the readings: photon counts per minute (CPM) at sensing node i due to source S, follow a Poisson distribution [8], [12] given by:

$$CPM_i = \frac{I_s \times A_i \times \eta_i}{(d_i^s)^2} \tag{6}$$

where CPM_i is the photons counts per minute at node i due to source with intensity I_s, A_i is the detector surface area, d_i^s is the distance between node i and the source S, and η_i is detector efficiency which is given by:

$$\eta_i = \frac{\#\ of\ photons\ recorded\ by\ i^{th}\ node}{\#\ of\ incident\ photons\ on\ i^{th}\ node} \tag{7}$$

Background radiation is negligible compared to the source radiation and, is considered same throughout the AoI. Thus, considering that all detectors are equally efficient and have same surface area A_i, we have:

$$CPM_i \propto \frac{I_s}{(d_i^s)^2} \tag{8}$$

While generating data, the source intensity of 10^9 has been taken into consideration. For anomalous behaviour simulation, three different seed values have been used to select three different sets of $n\%$ of nodes randomly. The readings of those randomly selected $n\%$ nodes are changed to random values in the range of 0 to 3 times the actual values. To report final evaluations at $n\%$ of anomalous nodes, the results have been averaged over three different sets.

Even though radiation source localization has been used for experiments, the proposed system is applicable for any localization task with a single target. This is due to the fact that data readings in localization tasks are a function of the distance of nodes from the target; the closer the nodes are to the target, higher the readings.

B. DATASET

To generate the data, nodes are synthetically placed within a square area in a uniform manner. The readings are simulated using Poisson's Distribution as mentioned in Section IV-A. Table 3 shows the structure of the dataset used in the experiments carried out.

TABLE 3. Dataset description.

Dataset Description
ID
Longitude
Latitude
Sensor Readings on different timestamp

C. EVALUATION RESULTS

Many experiments have been carried out to test the efficacy of our system. Since the objective of this work is to improve system efficiency without compromising on task accuracy, each experiment evaluates the system on metrics including Localization time (total time taken by the system) and Localization error (euclidean distance between actual source location and predicted source location). Using these metrics, the system's performance and scalability are assessed on three test cases: 1) Varying percentage of anomalous nodes, 2) Varying population size, and 3) Varying Area Length, which have been described in sections IV-C1, IV-C2, IV-C3 respectively. In each test case, the proposed system, EFLS, is compared to:

1) System 1: a synthetic system that performs the same localization algorithm over the whole AoI without eliminating anomalous nodes. This system is similar to the ones in [8] and [10].
2) System 2: a synthetic system that performs the same localization algorithm over the reduced AoI but without eliminating anomalous nodes.

This has been done to test the efficacy of reducing the AoI and that of eliminating anomalous nodes in the reduced AoI before performing localization. Since this work is concerned with the improvement in localization efficiency and accuracy with the use of fewer nodes of the new AoI versus use of all the nodes of the entire AoI, the same localization algorithm has been used in all the scenarios. This aids in illustrating that using nodes that are closer to the target is highly effective in increasing localization accuracy along with reducing computational cost. Hence comparison with other works has not been performed.

1) VARYING PERCENTAGE OF ANOMALOUS NODES

The 3 simulations aforementioned i.e: 1) System 1, 2) System 2, and 3) EFLS (performing localization over nodes in reduced area after eliminating anomalies), are tested by varying the percentage of anomalous nodes in the area between 0% to 50%. The system has been tested on varying percentages of anomalous nodes in order to evaluate its ability to handle different proportions of abnormal nodes. Evaluation over 50% anomalous nodes is not done because the majority of nodes start behaving randomly after that. This is due to the fact that the majority of the data points are anomalous and the system struggles to distinguish between the normal and the abnormal ones [16], [47]. Throughout the experiments, a fixed area size of 5 km by 5 km with 121 nodes is used.

Figure 6 describes the simulation results for localization error of the 3 systems for varying percentages of anomalous nodes. As can be seen, reducing only the area of interest (System 2) reduces the localization error to some extent, but the error deteriorates in the presence of higher anomalous nodes since most nodes within the reduced area also become anomalous. On the other hand, reduction of area in addition to removal of anomalies in EFLS reduces the error significantly by upto 68%. It is evident that EFLS is able to maintain a low and stabilized error for upto 30% of anomalous nodes. Likewise, Figure 7 describes the localization time taken by the 3 systems for varying percentages of anomalous nodes. However, unlike in the case of localization error, reducing the area (System 2) drastically reduces the time taken by the

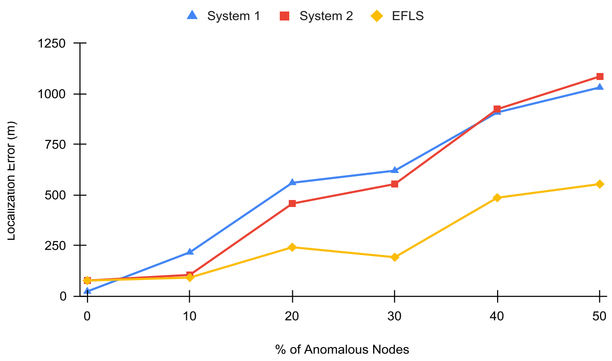


FIGURE 6. **The localization error for varying percentage of anomalous nodes.**

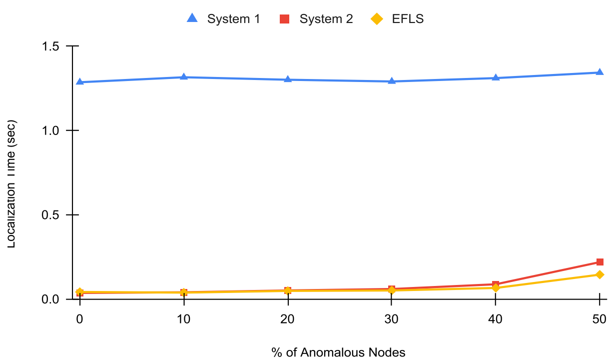


FIGURE 7. **The localization time for varying percentage of anomalous nodes.**

system to perform the localization in comparison to System 1, which is expected since there are fewer and more significant nodes in the reduced AoI. For upto 40% of anomalous nodes, EFLS takes similar time as System 2. However, for higher anomalous nodes, EFLS takes slightly less time. Thus, EFLS is able to reduce the time taken by the system from about 1.3 seconds to 30 milliseconds, which is around 30 times faster.

2) VARYING POPULATION SIZE

The same 3 systems mentioned in the section IV-C1 are tested for scalability by varying the number of nodes (population) in a fixed area of size 5 km by 5 km with 10% of anomalous nodes. The number of nodes are varied between 121 to 529. Figure 8 demonstrates that for different population sizes, EFLS shows significantly lower errors than System 1. Also it can be seen that while the reduction of area in System 2 improves localization error to some extent, the anomaly removal step in EFLS further reduces this error. Likewise, Figure 9 shows that the time taken by System 1 increases as the population size increases. However EFLS and System 2 are able to perform the task in a very low time for all population sizes. As such, EFLS is able to reduce the localization error by 79% to 95% and the localization time by 49 to 81 times. This proves the scalability of the proposed approach, EFLS, in terms of both accuracy and time for different population sizes.

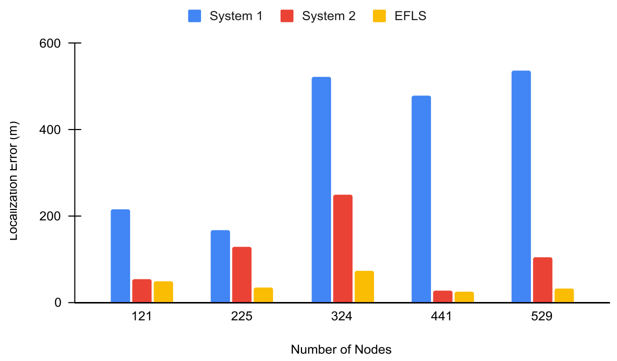


FIGURE 8. **The localization error for varying population size.**

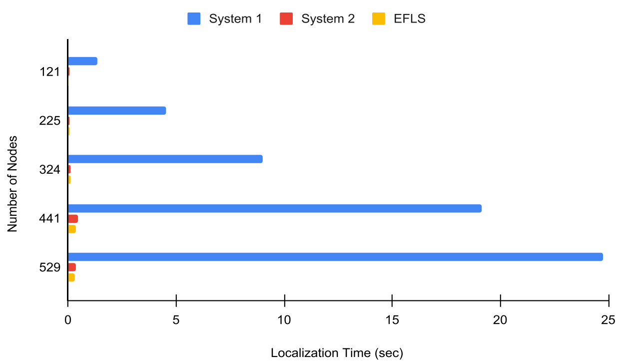


FIGURE 9. **The localization time for varying population size.**

3) VARYING AREA LENGTH

In this case as well, the 3 systems mentioned in the section IV-C1 are tested for scalability by varying the area size. The length of a square area with 225 nodes, of which 10% are anomalous, is varied between 1 km to 5 km. From Figure 10, it is evident that EFLS shows about 43% to 79% lower errors than both its counterparts for all area lengths. Additionally, EFLS and System 2 perform the localization task 52 to 65 times faster than the system using the whole area for all area lengths, this has been evident by Figure 11. This proves the scalability of EFLS in terms of both accuracy and time for different area lengths.

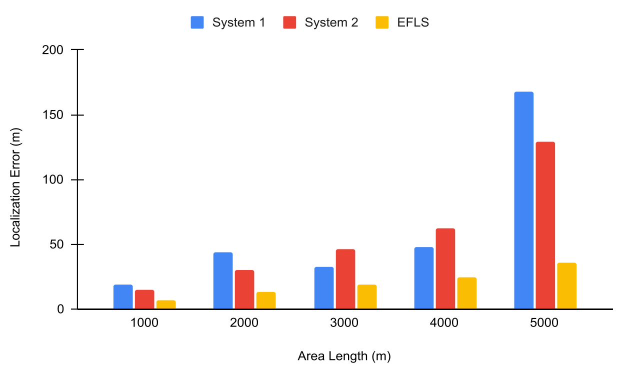


FIGURE 10. **The localization error for varying area length.**

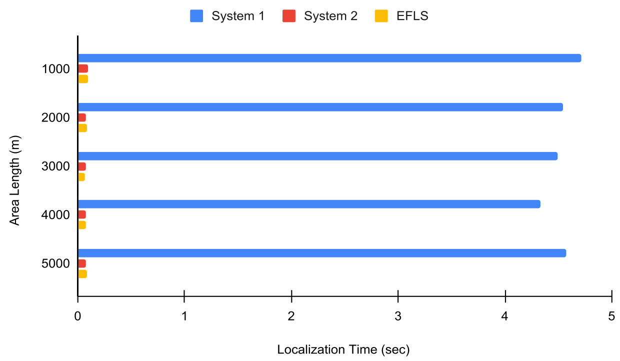


FIGURE 11. The localization time for varying area length.

V. CONCLUSION

In this paper, an efficient fault-proof localization system (EFLS) has been proposed for IoT and MCS. The system targets to reduce system complexity by providing an efficient mechanism of active node selection by narrowing the AoI to be around the target location in a way that is resilient to the presence of anomalous nodes. The system achieves this with a novel mechanism that employs the median image filtering technique to reduce the AoI. In addition, the system effectively detects and eliminates anomalies innovatively by comparing the actual and filtered values. In doing so, the system is able to overcome the shortcomings of computational complexities and inefficiency in existing works. The proposed system is tested on real-life dataset for radiation source localization tasks and compared to non-resilient systems of which one performs localization over the whole area while the other does the same in the reduced area. The results show that the proposed system reduces the localization error by up to 68% and localization time by up to 30 times for different percentages of anomalous nodes. The scalability of the proposed system in terms of population and area sizes is also verified. The results show that EFLS lessens the error by an average of 89% and 61%, respectively, and reduces the time required to perform the task significantly. Moreover the results demonstrate that the reduction of area in our system contributes to faster localization while anomaly removal in addition to area reduction is responsible for improved accuracy.

While the proposed system is able to improve the efficiency of localization significantly by integrating image filtering technique to address challenges of active node selection and anomaly handling, there is some room for further improvements. One is to detect anomalies over the whole area so as to penalise anomalous nodes and hence perform node reputation updates for future tasks. Following this, another direction for future work is taking into consideration other parameters such as residual energy, cost and reputation of nodes while selecting nodes in the reduced area. Additionally, the system currently only localizes a single stationary target. Further work can be done to expand it so as to localize multiple targets as well as mobile targets.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Ahmed Alagha for his valuable and constructive suggestions during the planning and development of this research work.

REFERENCES

[1] M. A. Feki, F. Kawsar, M. Boussard, and L. Trappeniers, "The Internet of Things: The next technological revolution," *Computer*, vol. 46, no. 2, pp. 24–25, Feb. 2013.

[2] D. Giusto, A. Iera, G. Morabito, and L. Atzori, *The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications*. New York, NY, USA: Springer-Verlag, Jan. 2010.

[3] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011.

[4] M. Kadadha, H. Otrok, R. Mizouni, S. Singh, and A. Ouali, "SenseChain: A blockchain-based crowdsensing framework for multiple requesters and multiple workers," *Future Gener. Comput. Syst.*, vol. 105, pp. 650–664, Apr. 2020.

[5] Z. Sisi and A. Souri, "Blockchain technology for energy-aware mobile crowd sensing approaches in Internet of Things," *Trans. Emerg. Telecommun. Technol.*, pp. 1–14, Jan. 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4217

[6] S. H. Marakkalage, S. Sarica, B. P. L. Lau, S. K. Viswanath, T. Balasubramaniam, C. Yuen, B. Yuen, J. Luo, and R. Nayak, "Understanding the lifestyle of older population: Mobile crowdsensing approach," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 1, pp. 82–95, Feb. 2019.

[7] J. Liu, H. Shen, H. S. Narman, W. Chung, and Z. Lin, "A survey of mobile crowdsensing techniques: A critical component for the Internet of Things," *ACM Trans. Cyber-Phys. Syst.*, vol. 2, no. 3, pp. 1–26, 2018.

[8] A. Liu, M. Wu, K. M. Chandy, D. Obenshain, M. Smith, and R. McLean. (2008). *Design Tradeoffs for Radiation Detection Sensor Networks*. [Online]. Available: http://www.cnds.jhu.edu/~dano/IPSN_final.pdf

[9] F. Asano, H. Asoh, and K. Nakadai, "Sound source localization in spatially colored noise using a hierarchical Bayesian model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 193–196.

[10] T. N. Le, P. H. J. Chong, X. J. Li, and W. Y. Leong, "A simple grid-based localization technique in wireless sensor networks for forest fire detection," in *Proc. 2nd Int. Conf. Commun. Softw. Netw.*, 2010, pp. 93–98.

[11] X. Luo, J. Yang, and L. Chai, "Water pollution source localization based on the contour in sensor networks," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2016, pp. 3373–3377.

[12] A. H. Liu, J. J. Bunn, and K. M. Chandy, "An analysis of data fusion for radiation detection and localization," in *Proc. 13th Int. Conf. Inf. Fusion*, Jul. 2010, pp. 1–8.

[13] Z. Liu, W. Dai, and M. Z. Win, "Node placement for localization networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[14] G. Popli and R. P. Singh, "Optimization of energy consumption in wireless sensor network—A review," in *Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2018, pp. 961–967.

[15] A. Alagha, S. Singh, R. Mizouni, A. Ouali, and H. Otrok, "Data-driven dynamic active node selection for event localization in IoT applications—A case study of radiation localization," *IEEE Access*, vol. 7, pp. 16168–16183, 2019.

[16] A. Alagha, S. Singh, H. Otrok, and R. Mizouni, "RFLS-resilient fault-proof localization system in IoT and crowd-based sensing applications," *J. Netw. Comput. Appl.*, vol. 170, Nov. 2020, Art. no. 102783.

[17] L. M. Kaplan, "Global node selection for localization in a distributed sensor network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 113–135, Jan. 2006.

[18] J. Feng, H. Zhao, and B. Lian, "Efficient and adaptive node selection for target tracking in wireless sensor network," *J. Sensors*, vol. 2016, Nov. 2015, Art. no. 9152962, doi: 10.1155/2016/9152962.

[19] A. Alagha, R. Mizouni, S. Singh, H. Otrok, and A. Ouali, "SDRS: A stable data-based recruitment system in IoT crowdsensing for localization tasks," *J. Netw. Comput. Appl.*, vol. 177, Mar. 2021, Art. no. 102968. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804520304197

[20] L. Chen, T. Chen, H. Chen, and M. Tsai, "Crowdsourced children monitoring and finding with holding up detection based on Internet-of-Things technologies," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12407–12417, 2019.

[21] S. K. Pandey and M. A. Zaveri, "Event localization in the Internet of Things environment," *Procedia Comput. Sci.*, vol. 115, pp. 691–698, Aug. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050917319865

[22] E.-W. Bai, K. Yosief, S. Dasgupta, and R. Mudumbai, "The maximum likelihood estimate for radiation source localization: Initializing an iterative search," in *Proc. 53rd IEEE Conf. Decis. Control*, Dec. 2014, pp. 277–282.

[23] J.-C. Chin, D. K. Y. Yau, N. S. V. Rao, Y. Yang, C. Y. T. Ma, and M. Shankar, "Accurate localization of low-level radioactive source under noise and measurement errors," in *Proc. 6th ACM Conf. Embedded Netw. sensor Syst. (SenSys)*, 2008, p. 183.

[24] F. Grondin, I. Sobieraj, M. Plumbley, and J. Glass, "Sound event localization and detection using CRNN on pairs of microphones," in *Proc. Detection Classification Acoustic Scenes Events Workshop (DCASE)*, 2019, p. 84.

[25] B. Lashkari, J. Rezazadeh, R. Farahbakhsh, and K. Sandrasegaran, "Crowdsourcing and sensing for indoor localization in IoT: A review," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2408–2434, Apr. 2019.

[26] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2nd Quart., 2010.

[27] C. Varun, B. Arindam, and K. Vipin, "Anomaly detection : A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[28] A. M. Hambal, Z. Pei, and F. L. Ishabailu, "Image noise reduction and filtering techniques," *Int. J. Sci. Res.*, vol. 6, no. 3, pp. 2033–2038, 2015.

[29] N. Win, K. Kyaw, T. Win, and P. Aung, "Image noise reduction using linear and non-linear filtering technique," *Int. J. Sci. Res. Publications*, vol. 9, no. 8, pp. 816–821, Aug. 2019.

[30] R. Chandel and G. Gupta, "Image filtering algorithms and techniques: A review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 10, pp. 198–202, Oct. 2013.

[31] M. Morelande, B. Ristic, and A. Gunatilaka, "Detection and parameter estimation of multiple radioactive sources," in *Proc. 10th Int. Conf. Inf. Fusion*, Jul. 2007, pp. 1–7.

[32] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2019.

[33] G. T. Whipps, E. Ertin, and R. L. Moses, "Distributed sensing for quickest change detection of point radiation sources," in *Proc. 18th Int. Conf. Inf. Fusion*, 2015, pp. 22–27.

[34] W. Wu, X. Cheng, M. Ding, K. Xing, F. Liu, and P. Deng, "Localized outlying and boundary data detection in sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1145–1157, Aug. 2007.

[35] L. Bettencourt, A. Hagberg, and L. Larkey, "Separating the wheat from the chaff: Practical anomaly detection schemes in ecological applications of distributed sensor networks," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst.*, Jun. 2007, pp. 223–239.

[36] M. C. Jun, H. Jeong, and C.-C. J. Kuo, "Distributed spatio-temporal outlier detection in sensor networks," *Proc. SPIE*, vol. 5819, pp. 273–284, Jun. 2005, doi: 10.1117/12.604764.

[37] J. Branch, B. Szymanski, C. Giannella, R. Wolff, and H. Kargupta, "In-network outlier detection in wireless sensor networks," in *Proc. 26th IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2006, p. 51.

[38] K. Zhang, S. Shi, H. Gao, and J. Li, "Unsupervised outlier detection in sensor networks using aggregation tree," in *Proc. Int. Conf. Adv. Data Mining Appl.*, Aug. 2007, pp. 158–169.

[39] H. Liu and S. Zhang, "Noisy data elimination using mutual k-nearest neighbor for classification mining," *J. Syst. Softw.*, vol. 85, no. 5, pp. 1067–1074, May 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121211003049

[40] B. Sheng, Q. Li, W. Mao, and W. Jin, "Outlier detection in sensor networks," in *Proc. 8th ACM Int. Symp. Mobile ad hoc Netw. Comput. (MobiHoc)*, 2007, p. 219, doi: 10.1145/1288107.1288137.

[41] M. Celik, F. Dadaser-Celik, and A. S. Dokuz, "Anomaly detection in temperature data using DBSCAN algorithm," in *Proc. Int. Symp. Innov. Intell. Syst. Appl.*, Jun. 2011, pp. 91–95.

[42] H. S. Emadi and S. M. Mazinani, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks," *Wireless Pers. Commun.*, vol. 98, no. 2, pp. 2025–2035, Jan. 2018, doi: 10.1007/s11277-017-4961-1.

[43] D. Janakiram, A. V. U. P. Kumar, and A. M. V. Reddy, "Outlier detection in wireless sensor networks using Bayesian belief networks," in *Proc. 1st Int. Conf. Commun. Syst. Softw. Middleware*, 2006, pp. 1–6.

[44] Y. Zhang, N. Meratnia, and P. J. M. Havinga, "Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine," *Ad Hoc Netw.*, vol. 11, no. 3, pp. 1062–1074, May 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870512002041

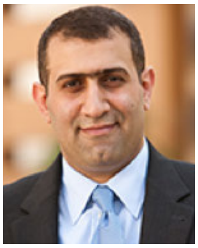
[45] D. J. Hill, B. S. Minsker, and E. Amir, "Real-time Bayesian anomaly detection in streaming environmental data," *Water Resour. Res.*, vol. 45, no. 4, pp. 1–16, Apr. 2009.

[46] A. Hameed and H. Mohammed, "Effect of different window size on median filter performance with variable noise densities," *Int. J. Comput. Appl.*, vol. 178, no. 2, pp. 22–27, Nov. 2017.

[47] V. Chatzigiannakis and S. Papavassiliou, "Diagnosing anomalies and identifying faulty nodes in sensor networks," *IEEE Sensors J.*, vol. 7, no. 5, pp. 637–645, May 2007.

ADARSH GHIMIRE received the B.E. degree in electronics and communication engineering from the Kathmandu Engineering College, Tribhuvan University, Kathmandu, Nepal. He is currently pursuing the M.Sc. degree in electrical and computer engineering with the Khalifa University of Science and Technology, Abu Dhabi, UAE. His research interests include computer vision, image processing, machine learning, generative adversarial networks, deep learning, robotics, and automations.

SELINA SHRESTHA received the B.E. degree in electronics and communication engineering from the Institute of Engineering, Tribhuvan University, Kathmandu, Nepal. She is currently pursuing the M.Sc. degree in electrical and computer engineering with the Khalifa University of Science and Technology, Abu Dhabi, UAE. Her research interests include data analytics, machine learning, and deep learning.



HADI OTROK (Senior Member, IEEE) received the Ph.D. degree in ECE from Concordia University, Montreal, QC, Canada. He holds an Associate Professor position with the Department of Electrical Engineering and Computer Science (EECS), Khalifa University. He is also an Affiliate Associate Professor with the Concordia Institute for Information Systems Engineering, Concordia University, and the Electrical Department, Ecole de Technologie Superieure (ETS), Montreal. His research interests include the domain of blockchain, reinforcement learning, crowd sensing and sourcing, ad hoc networks, and cloud security. He co-chaired several committees at various IEEE conferences. He is also an Associate Editor of IEEE Transactions on Network and Service Management (TSNM), *Ad-Hoc Networks* (Elsevier), and IEEE Network. He also served as an Associate Editor for IEEE Communications Letters from 2015 to 2019.