Event Detection Using the User Context in Sensor Based IoT

Anubhav Shivhare · Vishal Krishna Singh · Manish Kumar

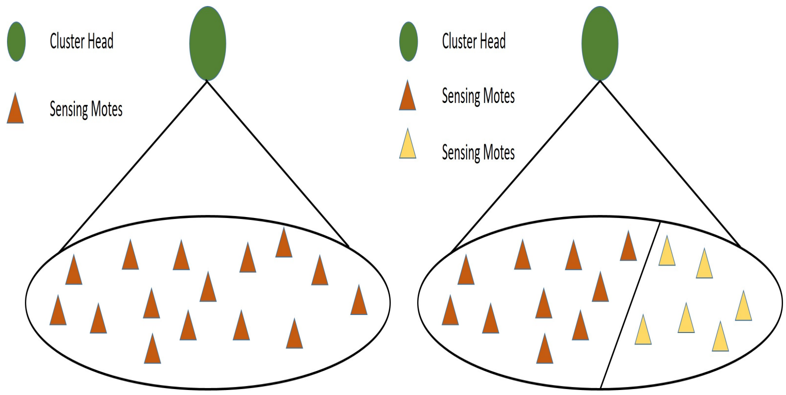
Abstract The clustering methodologies for deployed devices in the region of interest consider the similarity of sensed data or target maximization of network lifetime. Thus the clusters formed are either sensing clusters or communication clusters. In the case of motes capable of sensing multi-feature data, the similarity of sensed parameters is generally used to cluster the devices. Event detection methodologies in uni-modal sensor based IoT mainly focus on communication clusters to maximize network lifetime. Yet, the existing approaches for clustered data gathering hardly incorporate user context and preferences for sensing cluster formation and event detection. Additionally, many approaches do not incorporate both sensing and communication clusters in order to detect the event and balance it with energy efficiency. Research work that considers user context in the domain of the internet of things is also limited with a focus on network lifetime, event detection accuracy etc. The present work aims at resolving the existing constraints in 'user context' aware clustering methods and event detection. The user context parameters based on the domain knowledge of the user are used to divide the deployed region into sub-regions forming sensing clusters. The methodology allows the user to change the context and definition of events for each sub-region for accurate, context-aware, and discriminatory event detection. Additionally, the compressive gathering of detected events results in energy efficient data transmission. Simulations were performed for scalability and comparative analysis of the proposed scheme. The results show that the proposed scheme outperforms the existing schemes in terms of detection accuracy and network lifetime.

Keywords Event detection · sensor based IoT · clustering · user context · compressed sensing

Anubhav Shivhare, Manish Kumar
#4206, Computer Center-2, Indian Institute of Information Technology, Allahabad
Vishal Krishna Singh
Dept. of Information Technology, Indian Institute of Information Technology, Lucknow
E-mail: anubhavshivhare@gmail.com

1 Introduction

Sensor Based Internet of Things (SBIoT) is a smart and secure Machine to Machine (M2M) [1] network of sensor motes deployed in a Region of Interest (ROI). The SBIoT network deployment targets at ROI monitoring, data collection, and intelligent decision-making [2] [3]. Additionally, SBIoT should not only be scalable, fast and efficient but also sustainable in terms of energy efficiency and network lifetime [4] [5] [6]. SBIoT devices are motes with multiple sensors powered by an energy source, usually a battery. Additionally, motes may or may not be rechargeable depending upon the infrastructure available in ROI and accessibility of the deployed region. Billions of devices will be connected to a fully functional IoT network [2]. Hence, energy efficient algorithms for sensing, device clustering [7] and data transmission will reduce the overall energy budget of the network, increase the network lifetime and contribute to the objectives of green-IoT [8–10]. Moreover, IoT with increased bandwidth, data transmission motivates for incorporating user context while not compromising with the network lifetime [6] [11–13] in order to improve the quality of service to the user. Hence, a tradeoff between user context [14] aware methods and energy-sensitive approaches are required to balance discriminatory event detection and network lifetime. For example, Heating, Ventilation, and Air Cooling (HVAC) sys-



(a) Sensing and communication clusters are same (b) Sensing and communication clusters are different

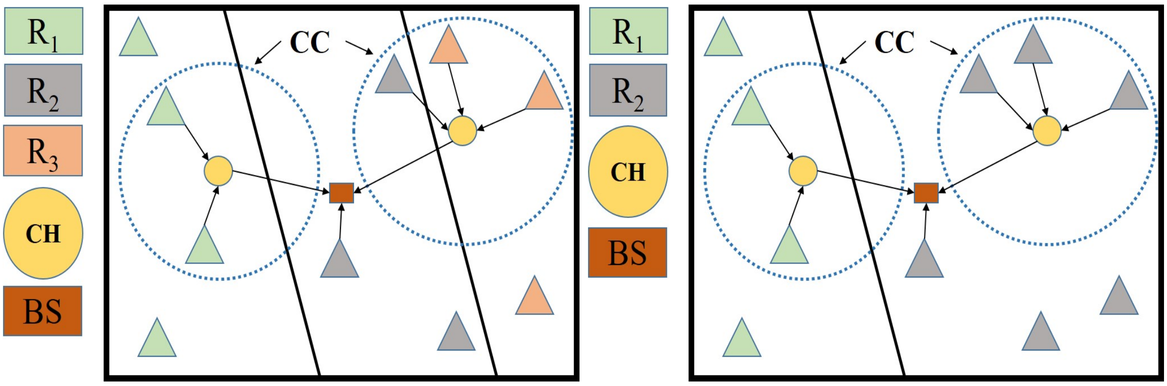
Fig. 1: Sensing and communication clusters

tems focus on maintaining user comfort while managing the energy budget in the deployed area [15–18]. Thus user-specific ambiance is to be maintained at different times and any deviation from it can be considered as an event for triggering the target action. Moreover, the event triggering definition has to change with change in user, seasonality, and application (industrial, home automation) etc. Hence accommodating the user context is required to update the event definitions in spatio-temporal domain.

Traditional schemes for clustered event detection are driven by one of the following two aspects i.e. (i) deployed devices are clustered based on the sensed data which results in the formation of *Sensing Clusters (SC)* and (ii) deployed devices are clustered to heed upon energy budget of the network which results in *Communication Clusters (CC)* (figure 1). Communication clusters primarily focus on energy efficient data transmission to Base Station (BS) with the aim of boosting network lifetime. Though, sensing clusters based on spatio-temporal data correlation may be totally different in comparison to communication clusters. Additionally, many approaches do not incorporate both approaches of clustering into account simultaneously energy efficient event detection while maintaining the sensing clusters. This phenomenon motivates sensed data dependent clustering approaches to place similar devices in the same clusters. Yet both approaches do not take user context awareness into account to create device-level clusters and detect user-specific events, respectively. Further, in case of multi-modal SBIoT multiple types of data can be sensed like temperature, humidity, smoke level etc. using the same mote. Hence composite event detection becomes complex and EOIs can not be accurately determined using one parameter only. Thus, user context based event detection becomes even more relevant in the case of composite event detection when sensed parameters are many and event definitions are mutable as per user context. Thus accurate event detection and minimization of false alarm rate becomes a new challenge.

Authors in [13] have shown that in reference to networks with multi-modal sensor devices, data dependent clustering algorithms like K-means and Fuzzy C-means outperform LEACH-C which is specifically designed for sensor based networks. Authors in [19] have designed a clustering based aggregation scheme for continuous data collection. The work aims to reduce the transmission cost. Spatio-temporal correlation of sensor based networks is exploited to calculate the trade-off between transmission of sensed data and prediction. Thus the majority of device clustering approaches fail to serve the purpose in upcoming IoT scenarios where motes are capable of complex computations [6]. In [20] authors argue that low energy consuming, low-latency and user-context aware protocols are inevitable. This becomes further necessary for efficient data transmission using existing peer-to-peer and wireless technologies. Additionally, much focus has not been given to develop user context aware algorithms for SBIoT network. Authors in [21] mention that one dimensional Compressive Sensing (CS) falls short of serving the purpose for spatio-temporally correlated networks. Hence the task of aggregation and data transmission is distributed between the aggregator mote and the neighboring mote for load distribution in the network. Further, CS models like Kronecker CS suffer from performance degradation with the increase in data dimensionality.

Thus there exists a lacunae of methodologies and algorithms that take user context into account while in-



(a) Day (b) Night

Fig. 2: Variation of the user context in ROI

corporating both SCs and CCs. The utility of such techniques is also emphasized in [6] for SBIoT networks. In [22] Dey claims that *the information to characterize the situation of an entity is context*. Further, *a system which delivers relevant information related to user's task is called context-aware system*. Similarly, *context modeling is useful in identifying concrete attributes of context which are usually user defined and evolve as per user needs*. Traditional context acquisition schemes are based on the following parameters:

1. 'Responsibility' of context acquisition i.e. data push (from sensor to software) or data pull (query based data collection from sensor).
2. 'Frequency' of context acquisition for instant events (like, security breach) and interval events (like, target tracking).
3. 'Source' of context acquisition methods can be direct (from sensor hardware) or indirect (computational derivation or user input based settings).

For example, ROI is a smart industrial area (say) equipped with HVAC system. The deployed motes in ROI have smoke, humidity, and temperature sensors, required to maintain specific conditions in its sub-regions (figure 2) during day and night. Further, Event of Interest (EOI) is any deviation from conditions required by the user in the sub-regions. In such a case, the similarity of the sensed data and user preferences can form the basis to divide ROI into sub-regions R_1, R_2, and R_3, respectively (figure 2). Figure 2a shows the uniformity of user context within sub-regions for event detection. During the day, R_1 has moist and cool conditions; R_2 has variable humidity and temperature and R_3 has warm and dry conditions. But during the night sub-regions R_2 and R_3 show similar sensor readings. Hence, R_2 and R_3 sub-regions should be clubbed to form a single sensing cluster while maintaining the erstwhile communication cluster during the night. Thus a sensor instance detected as an event in R_1 may be normal for R_3 hence no alarm should be raised for R_3. Event definitions may further change w.r.t day and night, seasonality and other evolving preferences of the user. Hence user domain knowledge is required to fragment the ROI for defining and detecting EOIs correctly despite spatial and temporal variations of sensed data in ROI. Moreover, other specific contributions of the paper are as follows:

1. In terms of device-level clustering of deployed motes, sensing, and communication clusters have been distinguished and a model is proposed to incorporate both of them.
2. The utility of user context has been emphasized for better service delivery to the user in the SBIoT domain.
3. A novel event detection scheme has been proposed that incorporates user context for defining and detecting events. Additionally, the model allows event definitions to be mutable and update the user context as it evolves further, for EOI detection.
4. User context aware and sensed data dependent clustering model is proposed to create dynamic sensing clusters in ROI. Yet communication clusters can be based on the energy budget of the deployed network.
5. A dynamic and adaptive device clustering model is designed so that sensing cluster membership changes with a change in spatio-temporal correlation and domain knowledge of the user.
6. Importance of decision making (detection) at the device level is emphasized to overcome drawbacks of multi-feature CS approaches and thus reduce the data transmission cost.

The paper is further structured as follows: Section 2 describes the problem addressed in the paper. Section 3 and section 4 provide the network model and methodology proposed in the present work. Further, an analysis of the proposed work is presented in section 5. Preliminaries for evaluation of the scheme are provided in section 6. Simulation & results are shown in section 7 and the work is concluded in section 8.

2 Problem Description

Clustering algorithms are mainly inspired by specific parameters like energy efficiency resulting in *Communication Clusters (CC)* or grouping of the devices based on sensed data resulting in *Sensing Clusters (SC)*. Such parameters fail to account for the deployed node-specific properties like the correlation between sensed data, location of the mote in ROI, user context, etc. for clustering the devices. Thus traditional clustering techniques have the inherent problem of unidirectional information flow (mote to sink). Though in practical scenarios the definition of the event is mutable in itself based on user context. This makes user interaction and context based clustering algorithms inevitable for SBIoT networks and involves bidirectional communication. Further, clustering algorithms overly focus on energy efficient transmissions of sensed data without considering spatio-temporal correlation of the motes and user context driven event detection that can reduce the amount of relevant data itself to be transferred. The case gains further importance in the case of composite (multi-feature data based) event detection which requires both SC and CC, in order to reduce the transferable data and energy efficient transmission of the data, respectively. Hence the user context aware and sensed data correlation based clustering can reduce the energy budget of the multi-modal SBIoT network. This is achieved by (a) User domain knowledge-based sensing clusters i.e. similar data is observed in a sensing cluster (b) each sub-region has its weight for features and event definition based on user context to reduce misdetection and (c) reducing transmission cost of heavy CS approaches for multi-modal data by detecting event at mote level itself and transmitting relevant data only from each communication cluster. Thus the discussed issues related to device level clustering can be summarized as follows:

1. User context aware clustering and event detection techniques are less researched and energy parameters are over-emphasized in the research literature for clustering the deployed motes.
2. Multi-sensor motes produce multi-modal data posing a high cost for data transmission within the network.
3. CS based approaches are computationally expensive in the case of multidimensional data gathering and transmission.

3 Network Model

The network in the proposed work is considered to be a connected graph $G(V, E)$ where V_i and V_j are vertexes and E_{ij} is the edge (communication link) between V_i and V_j such that $(i \neq j)$. A sensor mote V_i in the network has K sensors which sense different physical parameters (K is constant throughout the network for each mote). Vector $P_i =< P_{i1}, P_{i2}, ..., P_{ik} >$ such that $P_{ii'} (i = 1, 2, ..n)$ and $(i' = 1, 2, ..k)$ are the values of K features sensed by i^{th} mote i.e. V_i. Each mote V_i is equipped with an antenna for maintaining the communication between the mote to Cluster Head (CH) and CH to sink or Base Station (BS). Also, each mote is in communication radius R_c of its Cluster Head (CH). Other assumptions considered for the network are as follows:

1. User has prior knowledge about the ROI and shares his domain knowledge at the sink.
2. Region of interest (ROI) is fully covered in terms of sensing and communication coverage.
3. Each mote is location aware by using a localization algorithm like [23, 24].
4. Communication radius (R_c) is twice of Sensing radius (R_s) i.e. $R_c = 2 \times R_s$ while sink can communicate throughout the network.
5. If, node $V_i \in \eta(V_j) \iff V_j \in \eta(V_i)$, where $\eta(V_i)$ is clustered neighborhood of V_i.

4 Proposed Methodology

4.1 Cluster Formation

1. For the context U that user shares at sink, let $\{V_1, V_2, \ldots, V_n\}$ be the n active motes out of S motes deployed using uniform random distribution in ROI of $l \times l$ m^2 area.
2. Base station is aware of the location of active motes based on the location map using a localization algorithm (Section 3).
3. A Reference Vector (RV) is defined as a vector containing 'ideal' or normal values of each parameter in the region when no EOI has occurred in ROI. Thus for sub-region R_x in ROI $RV_x =< Q_{x1}, Q_{x2}, ..., Q_{xk} >$.
4. Such a vector can also be obtained for each sub-region using an initial mock run in the network such that, each active node V_i having K sensors transfers its sensed data to BS. Thus, BS has an initial data

Algorithm 1 Cluster Formation and Event Detection

Input :
l –ROI dimension $l \times l$
n –Number of active motes
(x_i, y_i), $i \in (1, n)$ Location information of active motes
D –Data matrix D (equation 1) of n active motes $\{V_1, V_2, ..., V_n\}$ for given context U
r –Number of sub-region identified by user R_x $(x = 1, 2, ..r)$ for given context U
RV_x –Reference vectors for each sub-region R_x associated with it for given context U
$w_{xi'}$ –Set of weights for each sub-region R_x associated with it for given context U

Output: Sensing clusters, communication clusters, and EOI detection based on user context.

procedure Sensing Cluster Formation
  Divide ROI into sub-regions R_x based on user context U i.e. shared domain knowledge or data matrix D received after the mock run.
  for Each mote $V_i(i = 1, 2, ..n)$ **do**
    for Each $R_x(x = 1, 2, ..r)$ **do**
      Assign the mote to a R_x based on its deployed location coordinates (x, y) and U.
      Assign weights set w_{ix} for mote belonging to sub-region R_x.
      Assign reference vector RV_x for mote belonging to sub-region R_x.
    end for
  end for
end procedure
while Network is Stable and user context is U **do**
  procedure Communication Cluster Formation .
    Use LEACH to create CC throughout the network.
    Select a set of motes $C = \{V_\alpha | \alpha$ is mote id of CH$\}$.
    Assign V_i to CH set such that maximum number of $\eta(V_i) > 0$ and $\eta(V_i) \neq \eta(V_j)$ if $(i \neq j)$.
  end procedure
  procedure Event Detection
    Algorithm (2)
  end procedure
end while

set of packets sent by nodes V_i (i=1 to n) each having K sensors. The received data at BS can be represented by the following data matrix D (equation 1). Such that, each column represents a feature or sensed entity like temperature, humidity, etc. and each row represents data values sensed by all the K sensors in mote V_i.

$$D = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1k} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nk} \end{pmatrix} \qquad (1)$$

5. User has prior knowledge of the ROI (either in form of domain knowledge or using a mock run) and shares the context U at the sink in form of
   – Number of sub-regions R_x $(x = 1, 2.., r)$ that should be formed in ROI based on the mock run data or user's domain knowledge.
   – Reference vectors $\{RV_1, RV_2, ...RV_r\}$ for each subregion R_x $(x = 1, 2.., r)$ in ROI such that $RV_x =< Q_{x1}, Q_{x2}, ..., Q_{xk} >$ such that, K is the number of features sensed by a deployed mote and Q_{xk} is the ideal value for K^{th} feature in sub-region R_x.
   – Set of weights for specific parameters $w_x =< w_{x1}, w_{x2}, .., w_{xk} >$ for each sub-region R_x of ROI. Weights are used to increase the importance of one or more features based on user context for discriminatory and sub-region specific event detection.
6. Thus steps 1 to 5 divide ROI to form sensing level clusters based on spatial (mote location), temporal (sensed data at a particular time instance) similarity, and user preferences U.
7. Further, communication clusters are formed by using state-of-the-art clustering algorithms that focus on enhancing network lifetime like LEACH or residual energy based algorithm proposed in [25] [26]. Based on two-hop communication model a set of motes $C = \{V_\alpha | \alpha$ is mote id of CH $\}$ is selected as cluster heads such that euclidean distance $|V_\alpha - BS| \leq R_c(\alpha = 1, 2, .., c)$ and $|V_\alpha - V_m| \leq R_c$ where V_m is a member mote in $\eta(V_\alpha)$. Thus a communication cluster may have its member motes from different sensing clusters depending upon the shared context and deployment in ROI. At the level of MAC layer, LEACH protocol is specifically used in this work as it adopts the principle of collision avoidance by creating the TDMA schedules during data transmissions [27, 28]. Thus it suits the needs of communication clusters. It is the baseline protocol hence, the network lifetime is not improved during simula-

tions by using advanced algorithms in comparison to LEACH.

8. After each round of data transmission step 7 is repeated until the user changes the context U or the first mote of the deployed network is dead. The major objective for choosing different communication clusters in each round is load balancing throughout the ROI (Algorithm 1).

Algorithm 2 Procedure for Event Detection

```
procedure EVENT DETECTION
    for Each sub-region R_x(x = 1, 2, ..r) do
        for Each mote V_i in R_x do
            Calculate P'_i and RV'_x using equations (2) and (3)
            Calculate e_ix using equations (4) and (5)
            if ((e_ix ∈ [0, 1]) & (e_ix < 1 − λ_x)) then
                Transmit value of 1 to V_α such that (V_α ∈ C) and (V_α ∈ η(V_i))
            else if ((e_ix ∈ [0, 1]) & (e_ix ≥ 1 − λ_x)) then
                Compute temp = |(|P_i| − |RV_x|)|
                if temp > thresh_mag_x then
                    Transmit value of 1 to V_α
                end if
            else
                Event not detected at V_i.
            end if
        end for
    end for
end procedure
```

4.2 Composite Event Detection

1. For each context $U = \{U_1, U_2, .., U_\theta\}$ shared by the user, each mote in sub-region R_x is assigned a reference vector RV_x $(x = 1, 2, ..r)$ based on its location information. Thus spatial correlation of deployed motes is taken into account for assigning a RV_x.
2. At each mote V_i in sub-region R_x vector $P'_i =< P'_{i1}, P'_{i2}, ...P'_{ik} >$ and $RV'_x =< Q'_{x1}, Q'_{x2}, ...Q'_{xk} >$ are computed such that

$$P'_{ii'} = \frac{P_{ii'}}{max(P_{ii'}, Q_{xk})} \tag{2}$$

and

$$Q'_{xk} = \frac{Q_{xk}}{max(P_{ii'}, Q_{xk})} \tag{3}$$

where $max(\cdot)$ returns maximum among the given arguments.

3. The weighted cosine $wcos(\cdot)$ of the normalized vectors P'_i and RV'_x is computed to find out the weighted cosine similarity parameter e_{ix} at i^{th} mote $V_i \in R_x$.
4. e_{ix} is defined as

$$\begin{aligned} e_{ix} &= wcos(P'_i, RV'_x) \\ &= \frac{\sum_{i'=1}^{k}(w_{xi'} \cdot P'_{ii'} \cdot Q'_{xi'})}{\sqrt{\sum_{i=1}^{k} w_{xi'} P'^{2}_{ii'}}\sqrt{\sum_{i=1}^{k} w_{xi'} Q'^{2}_{xi'}}} \end{aligned} \tag{4}$$

$$e_{ix} = \left\{ \frac{w_{x1} \cdot P'_{i1} \cdot Q'_{x1} + w_{x2} \cdot P'_{i2} \cdot Q'_{x2} + \cdots + w_{xk} \cdot P'_{ik} \cdot Q'_{xk}}{\sqrt{\sum_{i=1}^{k} w_{xi'} P'^{2}_{ii'}}\sqrt{\sum_{i=1}^{k} w_{xi'} Q'^{2}_{xi'}}} \right\} \tag{5}$$

such that $(i = 1, 2, ..n)$,$(i' = 1, 2, ..k)$ and $(x = 1, 2, ..r)$.

5. At each mote V_i, e_{ix} is computed with the help of RV_x assigned to it. If $e_{ix} \in [0, 1]$ and $e_{ix} < 1 - \lambda_x$ then EOI is said to have occurred and the value 1 is transferred to CH. Here, λ_x is the threshold value for allowed deviation from the corresponding sub-region's RV_x.
6. In case, $e_{ix} \in [0, 1]$ and $e_{ix} \geq 1-\lambda_x$ then P_i and RV_x are closely associated. Hence magnitude is taken into consideration i.e. $\frac{|(|P_i|-|RV_x|)|}{|RV_x|} \times 100 > thresh_{mag_x}$ implies that EOI has occurred and the value 1 is transferred to CH. Here $| \cdot |$ represents the magnitude of the vector and $thresh_{mag_x}$ is the threshold for the difference in magnitude of vectors in sub-region R_x.
7. In all other cases, EOI has not occurred in reference to the sensed data detected by mote V_i (Algorithm 2).
8. At CH all the values are collected from member motes to form an array B of received data i.e. 0s and 1s such that $|B| = |\eta(V_\alpha)|$ where, $|B|$ is the size of the array at CH and $V_\alpha \in C$. The array of received data is transmitted to BS using compressed sensing (Algorithm 3).

Schematic diagram of the proposed scheme is presented in figure 3 for easy understanding of the whole process.

5 Discussion and Analysis

5.1 Compressed Sensing for Data Transmission

Compressed sensing is one of the widely used approaches for gathering the sensor data. Let X be the original signal such that $(X \in R^N)$ in some basis Φ and having

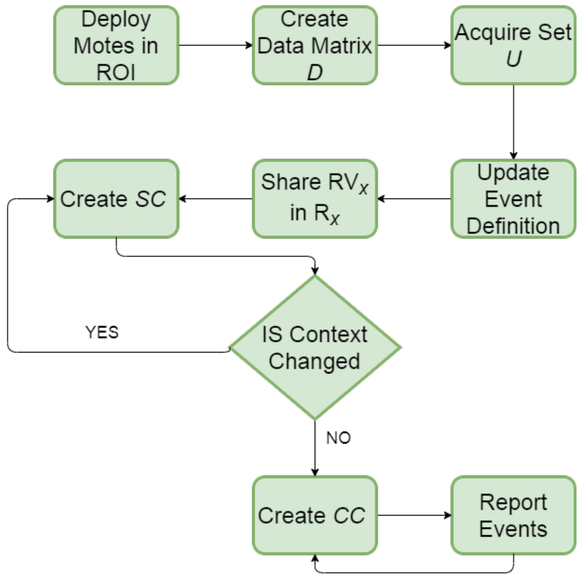


Fig. 3: Framework for user context aware event detection

sparsity (δ), a compressed signal $Y \in R^M(M << N)$ to be transmitted can be obtained (equation 6) such that,

$$Y = \Phi X \tag{6}$$

Sensor data is generally spatio-temporally correlated,

Algorithm 3 Compressed Sensing at CH

Input :
X –$X \in R^N$ is original sparse signal received at CH such that $X_n \in \{0, 1\}$.

Output:
Y –$Y \in R^M(M << N)$ is compressed signal to be transmitted from CH.

```
for Each instance of signal X do
    if (X_n = 1) then
        X_n ← mote_id
    end if
end for
procedure COMPRESSED SENSING
    Calculate a matrix Φ of order M × N using random global seed.
    Calculate Y ← Φ × X.
    Transmit Y to sink.
end procedure
```

hence it is sparse and compressible. Further, in the case of multi-modal data based composite event sensing, the cost of multi-dimensional compression is high [21]. Hence, the computation power of motes should be used for event detection at mote level instead of transmitting all the data to the sink level. To achieve this, user context sharing (in the form of domain knowledge) becomes important so that user context aware inferences can be derived from motes. Thus, data to be transmitted is reduced to 0 or 1 for each mote (Algorithm 1) i.e. $X_n = \{0, 1\}$ and problem of compression at CH again reduces to single feature compression. At CH, each instance of $X_n = 1$ is replaced by $mote_{id}$. Finally, the compressed signal $Y_{M\times1}$ is computed using $\Phi_{M\times N}$ and signal $X_{N\times1}$ which is transmitted to sink (Algorithm 3). The recovery is ensured by restricted isometric property. Additionally, a δ sparse signal can be efficiently compressed if equation (7) is satisfied.

$$M \leq c \cdot \delta \cdot log(N/\delta) << N \tag{7}$$

Further, the reconstructed signal is obtained by solving the following optimization problem:

$$\overline{X} = arg \min_{Y=\Phi\overline{X}} |\overline{X}|_1 \tag{8}$$

Equation (8), is the well-known l_1 optimization problem solved successfully for signal recovery in [29].

5.2 Weighted Cosine and Event Detection

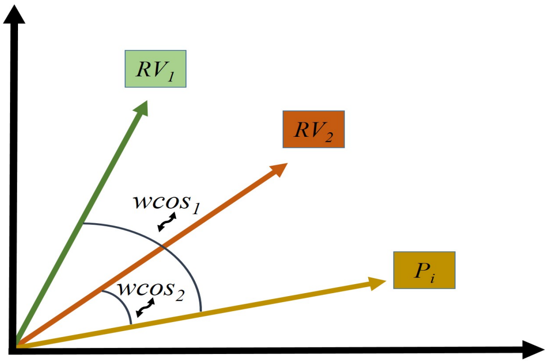
The cosine of two multi-feature vectors a and b is used to calculate the similarity between them (equation 9).

$$\text{cosine}(a, b) = \frac{\sum_{i=1}^{k} a_i b_i}{\sqrt{\sum_{i=1}^{k} a_i^2}\sqrt{\sum_{i=1}^{k} b_i^2}} \tag{9}$$

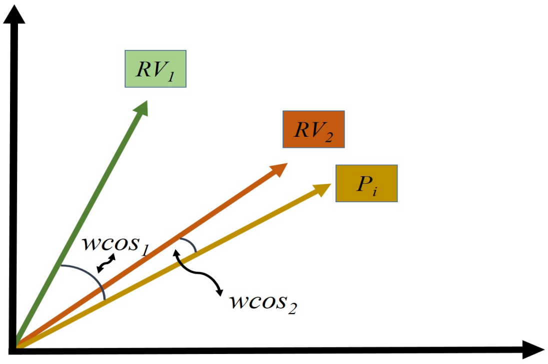
For K features, the memory and time complexity to compute equation (9) is $\mathcal{O}(\mathcal{K})$. Further, the vectors a and b are treated as the orthonormal basis in vector space. Thus, cosine helps in detecting the intensity of the event manifested as the angular distance between two vectors.

$$wcos(a, b) = \frac{\sum_{i=1}^{K} w_i a_i b_i}{\sqrt{\sum_{i=1}^{K} w_i a_i^2}\sqrt{\sum_{i=1}^{K} w_i b_i^2}} \tag{10}$$

Further weighted cosine (equation 10) allows the user to give additional weights to features based on the user's domain knowledge. The definition of EOI is itself mutable as per user needs and can be made biased towards particular features by assigning more weight to it for discriminatory event detection. For example, the user may assign more weight to temperature and smoke values in comparison to humidity for early detection of fire during winter. Moreover, EOIs can be exclusive to a region R_x depending upon RV_x (figure 4). If sensed P_i is dissimilar to RV_1 but similar to RV_2 then EOI is exclusively detected in R_1.



(a) High intensity (general) event



(b) Low intensity (exclusive) event

Fig. 4: Event detection in different sub-regions using RV_x and P_i

5.3 Optimal Cluster Size

Assuming the area of ROI to be A, number of active motes to be n, the deployment density per unit area is $\beta = \frac{n}{A}$. The coverage probability of a point in ROI can be calculated by using equation (11) by Koskinen [30].

$$P_{cov} = 1 - e^{-\beta \times \pi {R_s}^2} = 1 - e^{-\frac{(n_{CH}+n_m)\times \pi {R_s}^2}{A}} \tag{11}$$

such that n_{CH} is the number of cluster heads and n_m is the number of member motes. Further, the probability that a mote can directly reach the CH can be calculated using equation (12).

$$P_{conn} = 1 - e^{-\frac{n_{CH}\times \pi {R_c}^2}{A}} \tag{12}$$

Hence the total probability for a point to be covered and connected can be given by equation (13).

$$P_{conn} = 1 - e^{-\frac{n_{CH}\times A(CH)}{A}} \tag{13}$$

where $A(CH)$ is the expected value of the area covered by each cluster (including the CH and member motes). Further n_{avg}, the expected average number of member motes in a cluster can be found using equation (14)

$$n_{avg} = \frac{n_m}{n_{CH}}\left(1 - e^{-\frac{n_{CH}\pi R_c^2}{A}}\right) \tag{14}$$

The number of motes required α in ROI so that each CH has n_{avg} motes can be calculated as per equation (15).

$$\frac{n_{avg}}{\alpha} = \frac{\pi R_c^2}{A} \Longleftrightarrow A = \frac{\alpha}{n_{avg}/\pi R_c^2} \Longleftrightarrow \alpha = \frac{A \cdot n_{avg}}{\pi R_c^2} \tag{15}$$

For a point, P to be covered by a CH and its member motes, at least one mote should be present in area $G(z)$

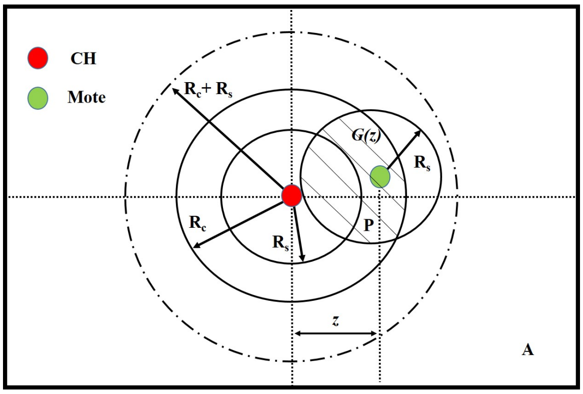


Fig. 5: Area covered by a member mote and cluster head

which is a function of the distance between the center (CH) and connected mote (figure 5).

$$G(z) = R_s^2 \cos^{-1}\left(\frac{z^2+R_s^2-R_c^2}{2zR_s}\right) + R_c^2 \cos^{-1}\left(\frac{z^2+R_c^2-R_s^2}{2zR_c}\right) - \frac{1}{2} \sqrt{(R_s+R_c-z)(z+R_s-R_c)(z-R_s+R_c)(z+R_s+R_c)} \tag{16}$$

$G(z)$ is calculated using equation(16) when $R_c - R_s < z \leq R_c + R_s$. Equation (17) gives $G(z)$ if $z \leq R_c - R_s$.

$$G(z) = \pi R_s^2 \tag{17}$$

The probability $Pr(z)$ that a point P is z distance away from a CH and is sensed by at least one connected sensor is given by equation (18).

$$Pr(z) = 1 - e^{-\frac{G(z) n_{avg}}{\pi R_c^2}} \tag{18}$$

Thus the total expected area covered by each cluster $A(CH)$ can be given by (as in [31]) using the cylindrical coordinates (equation 19).

$$A(CH) = \int_{z=R_s}^{R_s+R_c} \int_{\phi=0}^{2\pi} z\left(1-e^{-\frac{G(z)n_m}{\pi R_c^2}}\right) d\phi dz \\ = 2\pi \int_{z=R_s}^{R_s+R_c} z\left(1-e^{-\frac{G(z)n_m}{\pi R_c^2}}\right) dz \tag{19}$$

Rearrangement of terms gives us equation (20)

$$A(CH) = \pi R_s^2 + 2\pi \int_{z=R_s}^{R_s+R_c} z\left(1-e^{-\frac{G(z)n_m}{\pi R_c^2}}\right) dz \tag{20}$$

By solving the integral the final value of $A(CH)$ can be computed as per equation (21) for which $G(z)$ is approximated as per equation (22)

$$A(CH) = \pi R_s^2 + \\ 2\pi \int_{z=R_s}^{R_s+R_c} z\left(1-e^{-\frac{\left[-\left(\frac{\pi R_s}{2}\right)z+\left(\frac{\pi R_s}{2}\right)(R_s+R_c)\right]n_m}{\pi R_c^2}}\right) dz \tag{21}$$

$$G(z) = -(\pi R_s/2)\, z + (\pi R_s/2)\,(R_s+R_c) \tag{22}$$

Considering the case of $R_t \leq 2 \times R_t$ the area covered by a CH and its member motes is given by equation (23).

$$A(CH) = \pi\,(R_s+R_c)^2 + \\ \frac{2\pi}{\gamma}\left[\left(\frac{1}{\gamma}-R_s\right)\left(1-e^{-\gamma R_c}\right)-R_c\right] \tag{23}$$

where $\gamma = \frac{n_m R_s}{2R_c^2}$.

6 Preliminaries for Evaluation Metrics

6.1 Metrics of Evaluation

The proposed scheme is evaluated using three metrices i.e. (a) complexity analysis (b) accuracy of event detection and (c) network lifetime. Computation and communication cost both have been analyzed in complexity analysis. Further, Event Detection Accuracy (EDA) is used as a conventional accuracy metric to find the efficiency of any classification algorithm and can be computed using equation (24)

$$EDA = \frac{TP+TN}{TP+FP+TN+FN} \tag{24}$$

Similarly, False Alarm Rate (FAR) is calculated using equation (25).

$$FAR = \frac{FP}{FP+TN} \tag{25}$$

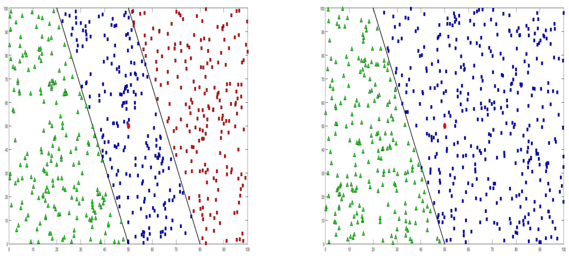
Additionally, network lifetime has been compared in terms of the overall energy budget and average residual energy of the deployed network.

6.2 Communication and Computation Cost Analysis

The proposed methodology consists of three major parts (a) sensing and communication cluster formation (b) event detection (c) compressed sensing for data communication. Sensing cluster formation requires $RV_x(x = 1, 2, .., r)$ assignment to each mote based on its location. Further communication clusters allow single-hop communication with CH which transmits data to the sink. Thus the sensing and communication complexity is of order $\mathcal{O}(\frac{N}{J})$ where n is the number of motes in the sensing field and J is the number of motes sensing the EOI. Additionally, event inference at leaf motes requires cosine computations having order $\mathcal{O}(K)$(section 5.2) and CS at CH level is bounded by order of $\mathcal{O}(M \times (\frac{N}{J})^2)$ (section 5.1).

6.3 Dataset

As per the premises of the proposed scheme, two types of user contexts $\{U_1, U_2\}$ were considered for day and night, respectively. U_1 (figure 6a) divides ROI into three sub-regions $R_x(x = 1, 2, 3)$ of 35% , 30% and 35%. Similarly U_2 divides ROI into two sub-regions $R_x(x = 1, 2)$ of 35% and 65% (figure 6b). To create RV_x for each



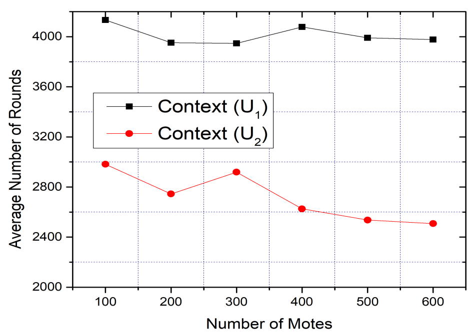
(a) User context U_1 (b) User context U_2

Fig. 6: Division of ROI based on user context

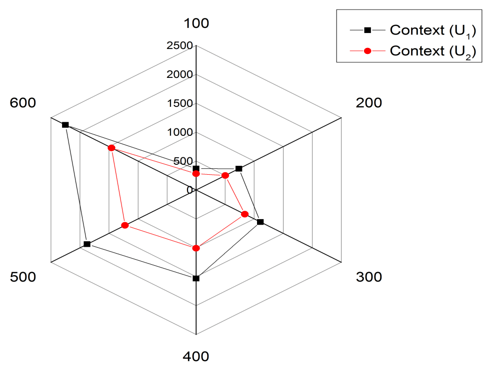
R_x humidity, smoke, and temperature attribute values were considered from dataset [32] collected using three different devices placed in three different surroundings. The devices used for data collection are aimed at maintaining (a) cool and humid (stable condition) in R_1 (b) variable conditions of temperature and humidity in R_2 and (c) warm and dry condition (stable condition) in R_3. U_1 considers all three conditions (a, b, and c), and U_2 considers only two conditions(a and c). For user context aware event detection the magnitude of allowed deviation ($\lambda = 0.3$ & $thresh_{mag} = 10\%$) from the ideal conditions was considered for event detection in all the sub-regions. Additionally $wcos(\cdot)$ was computed based on section 4.2.

7 Simulation and Results

7.1 Scalability Analysis



(a) Network lifetime



(b) Total number of packet transmissions (in thousands)

Fig. 7: Scalability of proposed scheme

Simulation Parameters	Value(s)
ROI dimensions	100×100 m^2
Number of motes deployed	100 to 600
Initial energy	0.5 J
Transmission energy	50 nJ

Table 1: List of simulation parameters for scalability analysis

For scalability analysis an ROI of 100×100 m^2 was considered with varying numbers of deployed motes at $R_s = 10$ m. and $R_c = 2 \times R_s$. Other parameters are summarized in Table 1. User context $U = \{U_1, U_2\}$ was considered separately and the stability period (before the first mote dead) was considered to compare the network lifetime. A mote was considered dead if mote energy is less than 1% of initial energy. Further random EOIs were generated throughout the ROI. Though EOIs were detected exclusively by a single $R_x (x = 1, 2, 3)$ based on $wcos(\cdot)$ in each data collection round. This was specifically done to avoid preferential treatment of a particular sub-region over others due to deployment randomness. Figure 7, shows the comparison of U_1 and U_2 in terms of network lifetime and in-network transmission. The following reasons are associated with the observed network lifetime trend:

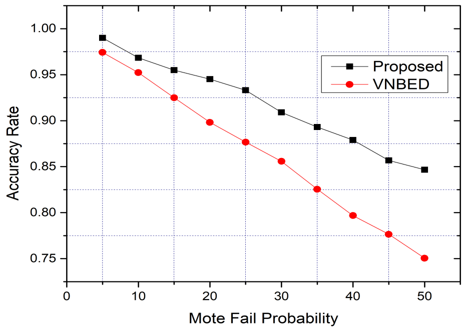
1. Average area coverage ratio is 94.63% at 100 motes and 99.5% at the deployment of more than 200 motes. Hence the number of rounds decreases figure(7a) from 100 to 200 for both U_1 and U_2 due to an increase in the in-network data transmission (7b).
2. Once sufficient area coverage is achieved (after 300 motes), the network lifetime achieves a maximum and becomes stable despite an increase in deployed motes. This is because all motes detecting a single EOI transmit data to CH instead of one mote. Hence, after a sufficient deployment of active motes, the network lifetime does not increase with an increase in deployment density.

7.2 Comparative Analysis

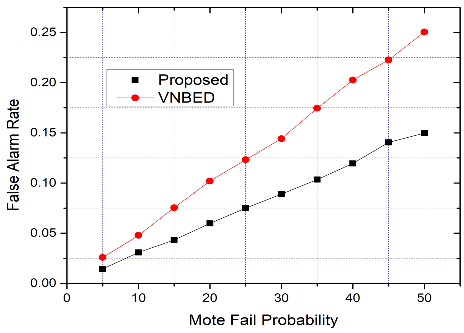
Simulation Parameters	Value(s)
ROI dimensions	200×200 m^2
Initial energy	0.5 J
Transmission energy	50 nJ

Table 2: List of simulation parameters for comparative analysis

For comparative analysis, simulations were performed by considering an area of 200×200 m^2. Based on the user context awareness the whole ROI was divided into R_1, R_2, and R_3 using U_1. Additionally, R_1 and R_2 were considered for U_2. EOIs were randomly generated throughout the sensing field. Based on the user context $U = \{U_1, U_2\}$, sub-region (R_x) and reference vector (RV_x) EOIs were detected by motes. The proposed scheme was evaluated based on parameters of the accuracy of EOI detection and network lifetime. Rest of the parameters are summarized in Table 2.



(a) Accuracy



(b) False alarm rate

Fig. 8: Performance of the proposed scheme in terms of accuracy

Accuracy of Event Detection For finding event detection accuracy 300 motes were deployed in 200 $\times$ 200 m^2 with a sensing radius 15 m to average out the simulation parameters in [33]. Figures (8a) & (8b) show the variation of mote fail probability with the accuracy of event detection and false alarm rate, respectively (Section 6.1). EOIs were randomly generated throughout the ROI. Deployed motes were further assigned with mote failure probabilities (0.05 to 0.5) throughout ROI to calculate the accuracy of detection. The results were further averaged out over 200 data collection rounds to eliminate the impact of abrupt outcomes. It is clear from the figure (8) that the decrease in accurate detection is gradual proposed scheme outperforms VNBED [33] by 6.60% & 67.01% in terms of accuracy rate and false alarm rate, respectively.

Network Lifetime Network lifetime is an important parameter to analyze the performance of any algorithm in terms of energy efficiency. For comparative analysis of network lifetime 200 motes were randomly deployed in 200 $\times$ 200 m^2 with R_s = 20 m. Further random number of EOIs were generated throughout the field.

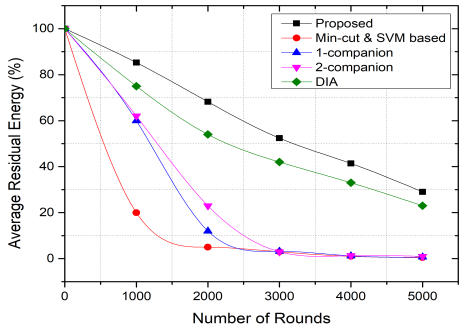


Fig. 9: Average residual energy varying with the number of rounds

Moreover, user context was considered to divide ROI into sub-regions R_1, R_2 & R_3 each covering 35%, 30% & 35% area, respectively. To compare the performance of the proposed scheme with [29] and [34], 5000 data collection round were considered. Events were detected exclusively to each region in cycles of R_1, R_2 & R_3. For calculating the average residual energy, the following energy dissipation model was used for transmitting (equation 26) and receiving (equation 27) l-bits of message [35]:

$$E_{\mathrm{Tx}}(l,d) = \begin{cases} l e_{\mathrm{elec}} + l\varepsilon_{fs} \cdot d^2 & \text{if, } d < d_o \\ l e_{\mathrm{elec}} + l\varepsilon_{\mathrm{amp}} \cdot d^4 & \text{if, } d \geq d_o \end{cases} \tag{26}$$

and

$$e_{Rx}(l) = l \cdot e_{\mathrm{elec}} \tag{27}$$

such that, $e_{\mathrm{elec}} = 50nJ/$ bit, $\varepsilon_{fs} = 10$ pJ/bit/m^2, $\varepsilon_{amp} = 0.0013$ pJ/bit/m^4. It is clear from the figure (9) that the proposed scheme outperforms DIA [34] in 5000 data collection rounds. The major reason for the proposed methodology to perform better in terms of network lifetime is discriminatory event detection. An event is considered exclusive to R_1 if it is detected in R_1 but ignored in R_2 and R_3. Thus exclusive events to a single sub-region were detected in each data collection round (section 4.2). This was done exclusively to take user context-aware event definition into account. Hence, the proposed methodology saves a major part of the deployed network's energy budget while 90% of the nodes are dead within 5000 rounds for the compared methods.

8 Conclusion

The problem of device level clustering and event detection in SBIoT framework has been studied in depth and a novel scheme is proposed to take user context into account for creating sensing clusters and detecting EOIs. Further, the scheme was made energy efficient using communication clusters and compressive data gathering. A detailed explanation and analysis of the proposed methodology are provided. Simulations were performed to test the efficacy of the scheme on various parameters. Scalability analysis and comparison with other schemes show that the proposed scheme outperforms the existing schemes (VNBED) in terms of detection accuracy. Similarly, schemes like Mincut & SVM based, 1-companion, 2-companion, and DIA are outperformed by the proposed scheme in terms of network lifetime. Thus, the proposed scheme achieves the objective of creating sensing clusters and event detection despite event definition being mutable in sub-regions based on user context. In the future, a combination of user context with other existing area coverage, clustering and data aggregation protocols in SBIoT can be explored to further improve user-specific event detection.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest with any financial or personal entity whose interests could be positively or negatively influenced by the article's content.

Availability of data and materials

No private dataset is associated with the present paper. The link for the publicly available dataset is provided in [32].

References

1. P.K. Malik, R. Sharma, R. Singh, A. Gehlot, S.C. Satapathy, W.S. Alnumay, D. Pelusi, U. Ghosh, J. Nayak, Computer Communications **166**, 125 (2021)
2. S. Li, L. Da Xu, S. Zhao, Journal of Industrial Information Integration **10**, 1 (2018)
3. W. Chen, Computer communications **151**, 31 (2020)
4. A. Shukla, S. Tripathi, Wireless Networks **26**, 3471 (2020)
5. M. Lu, G. Fu, N.B. Osman, U. Konbr, Sustainable Cities and Society **75**, 103349 (2021)
6. L. Xu, R. Collier, G.M. O'Hare, IEEE Internet of Things Journal **4**(5), 1229 (2017)
7. S. Ebrahimi Mood, M.M. Javidi, Evolving systems **11**(4), 575 (2020)
8. A. Shivhare, V.K. Singh, M. Kumar, IEEE Systems Journal **14**(4), 4854 (2020)
9. M.A. Albreem, A.M. Sheikh, M.H. Alsharif, M. Jusoh, M.N.M. Yasin, IEEE Access **9**, 38833 (2021)
10. A. Mukherjee, P. Goswami, L. Yang, Z. Yan, M. Daneshmand, Computer Communications **152**, 119 (2020)
11. K. Wallis, F. Schillinger, E. Backmund, C. Reich, C. Schindelhauer, in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)* (IEEE, 2020), pp. 7–14
12. C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, IEEE communications surveys & tutorials **16**(1), 414 (2013)
13. F. Medhat, R.A. Ramadan, I. Talkhan, in *2012 Seventh International Conference on Broadband, Wireless Computing, Communication and Applications* (IEEE, 2012), pp. 367–372
14. Z.A. Almusaylim, N. Zaman, Wireless networks **25**, 3193 (2019)
15. H. Xu, Y. He, X. Sun, J. He, Q. Xu, Computer Communications **151**, 581 (2020)
16. L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, IEEE Internet of Things Journal **7**(4), 2751 (2019)
17. G. Gao, J. Li, Y. Wen, IEEE Internet of Things Journal **7**(9), 8472 (2020)
18. H. Ping, Z. Zhou, Z. Shi, T. Rahman, Personal and Ubiquitous Computing **22**(3), 597 (2018)
19. H. Jiang, S. Jin, C. Wang, IEEE Transactions on Parallel and Distributed Systems **22**(6), 1064 (2010)
20. H. Joh, I. Yang, I. Ryoo, Peer-to-Peer Networking and Applications **9**(3), 520 (2016)
21. L. Quan, S. Xiao, X. Xue, C. Lu, IEEE Communications Letters **20**(3), 578 (2016)
22. A.K. Dey, G.D. Abowd, D. Salber, Human–Computer Interaction **16**(2-4), 97 (2001)
23. A.P. Behera, A. Singh, S. Verma, M. Kumar, IEEE Sensors Letters **4**(10), 1 (2020)
24. Y. Xiuwu, L. Ying, L. Yong, Y. Hao, Wireless Networks **27**(4), 3053 (2021)
25. T.M. Behera, S.K. Mohapatra, U.C. Samal, M.S. Khan, M. Daneshmand, A.H. Gandomi, IEEE Internet of Things Journal **6**(3), 5132 (2019)
26. A.O.A. Salem, N. Shudifat, Personal and Ubiquitous Computing **23**(5), 901 (2019)
27. K. Shaukat, F. Iqbal, I.A. Hameed, M.U. Hassan, S. Luo, R. Hassan, A. Younas, S. Ali, G. Adeem, A. Rubab, et al., in *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (IEEE, 2020), pp. 421–426
28. R.I. Tandel, International Journal of Computer Science and Information Technologies **7**(4), 1894 (2016)
29. C.H. Liu, J. Zhao, H. Zhang, S. Guo, K.K. Leung, J. Crowcroft, IEEE Systems Journal **11**(4), 2490 (2016)
30. H. Koskinen, in *Proceedings of 16th ITC specialist seminar* (Citeseer, 2004), pp. 11–18
31. C. Sevgi, A. Koçyigit, IEEE Communications Letters **12**(4), 232 (2008)
32. G.A. Stafford. Environmental sensor telemetry data (2020)
33. V.K. Singh, R. Verma, M. Kumar, AEU-International Journal of Electronics and Communications **82**, 358 (2017)
34. V.K. Singh, M. Kumar, S. Verma, IEEE Systems Journal **13**(1), 248 (2017)
35. Y. Li, N. Yu, W. Zhang, W. Zhao, X. You, M. Daneshmand, in *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)* (IEEE, 2011), pp. 223–228