# Leveraging ambient sensing for the estimation of curiosity-driven human crowd


Anirban Das
*Department of CSE*
*IIIT Guwahati, India*
anirban@iiitg.ac.in

Kartik Narayan
*Department of CSE*
*IIT Jodhpur, India*
narayan.2@iitj.ac.in

Suchetana Chakraborty
*Department of CSE*
*IIT Jodhpur, India*
suchetana@iitj.ac.in



***Abstract*—Identification and characterization of human crowd formulation have been a topic of immense interest in recent times due to its applicability in a wide range of smart-city applications covering infrastructure automation to targeted advertising. The core idea is to extract the dynamics and associated behavioural patterns of mass gatherings within an environment through a continuous remote monitoring of the crowd. In general, the existing approaches heavily rely on computer vision and image processing based algorithmic tools and techniques to address this problem or mandate the crowd entities to carry a smartphone with them. However, considering the ubiquitous design goals of futuristic smart applications, camera and smartphone driven *active* sensing is not suitable to honour users' right to privacy by requiring an active user participation. In this work, we introduce a novel approach towards measuring the spatio-temporal significance of an object in terms of the curious crowd it has attracted over the others. The proposed approach utilizes a set of passive sensors and Wireless signal properties for the necessary estimation. We validate the idea using a room-scale testbed with rigorous experimentation in a real-world scenario. The low cost solution has minimal invasive footprints towards privacy and is capable to reach beyond 90% of accuracy for this measurement.**



***Index Terms*—Crowd, IoT, Ambient, Sensing, CSI, Passive**


## I. Introduction

Over the past decade, the study of human occupancy has reached a new height. Identification of human gathering [1] and characterization of the crowd-level behavior [2] are generally part of the study that are of interest for many intelligent applications in the domain of home automation, business and marketing, assisted and healthy living, tourism and transportation to name a few. For instance, a conventional *hypermarket* could be interested in tracking which of it's sections or products are observing significant gatherings at certain times of a day. Usually tracked in terms of *retail bounce rate*, this information could help in providing adequate retail offers to woo the customers in real-time. Similarly, an art-gallery may wish to track, which of its art-works have attracted a significant amount of audience rather than the passersby. *Curiosity* is an innate quality of human and plays an important role in individual's behavior [2]. It can be considered as a naturally instilled factor to drive the crowd dynamics. An object or an event that triggers curiosity (in psychological terms) draw crowds towards it which have inherent temporal characteristics in terms of quantity as well as dynamics. For instance, a curious set of individuals can form smaller groups in front of a notice board relevant to their own interest (shown in Fig. 1) or engage in break-time discussions centered around tables (shown in Fig. 2). It is worth noting that curiosity-driven crowd dynamics, i.e. the formation and dissolution of the gatherings, is not quite persistent over time. For instance, Fig. 3 shows the change in the size of gathering over a duration of 2 hours within an area of our experimental environment.

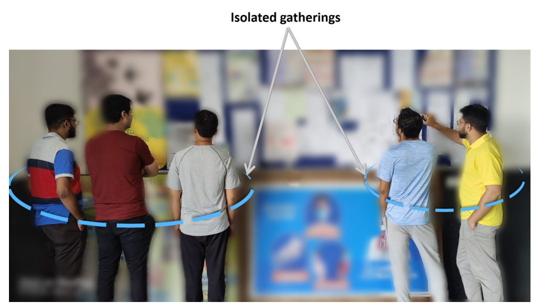



Fig. 1: Visibly isolated groups in front of notice boards

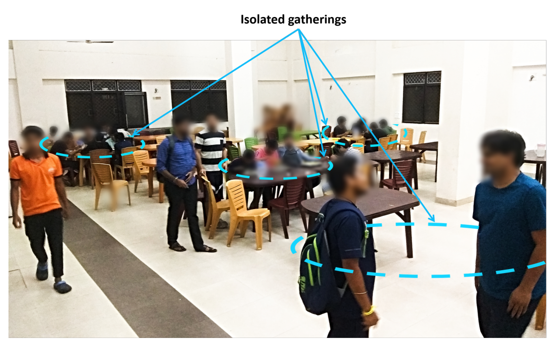



Fig. 2: Gathering of individuals during break-time

Active sensing approaches that use images or video feeds suffer from a significant challenge while addressing the privacy of user data [3]. Moreover, quality of the images, deployment restrictions, high computational overhead, high installation and maintenance cost, etc. are some of the major limitations that hinder the process of adopting these techniques for ubiquitous smart infrastructure. Similar concerns also apply for environment sensing that uses microphones and other identity exposing data sources. On the other hand, the approaches

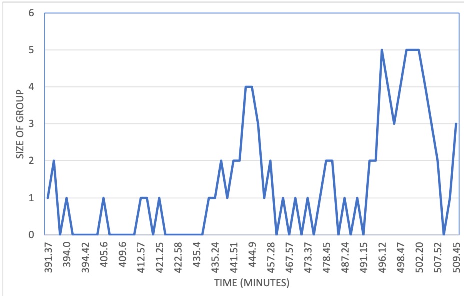



Fig. 3: Gathering dynamics over a duration of 2 hours.

which require installation of applications in the users' devices or sharing of personal information automatically questions the achievable ubiquity. Towards this, *passive sensing* techniques, free from active user intervention, are much effective to capture the crowd dynamics seamlessly. With the advent in MEMS technology, more reliable and cost-effective sensor devices gained significant attention these days to sense physical phenomena like motion, temperature, humidity, illuminance, $CO_2$, etc. which in turn could be correlated to understand the crowd level behavior. Similarly, characteristics of radio waves could potentially be leveraged to detect the presence of obstacles in its propagation path. Received Signal Strength Indicator (RSSI) [4] and Channel State Information (CSI) [5] have been explored in recent times to study human occupancy. However, the domain still faces a lot of challenges in terms of hardware compatibility, availability of supporting firmware and limitations of the achieved occupancy count [6] and this domain still has significant scopes.

That said, this work aims to *passively monitor curiosity-driven small-sized human gatherings around an object of interest in an indoor environment in real-time*, rather than targeting the overall room occupancy measures, as studied by the previous works [7], [8]. In this regard, the following questions are addressed: 1) Is it possible to identify which among an available set of objects triggered more interest among the individuals in terms of crowd being attracted? 2) Is it possible to estimate the size of such individual gatherings? 3) Is it possible to achieve these goals with minimal invasive profile? The key contributions of this paper can be stated as follows: 1) A low-cost minimally intrusive approach to estimate curiosity-driven gatherings in an indoor environment 2) Multi-sensor and CSI fusion for estimating gatherings and their dynamics in a passive manner. 3) A real test-bed based evaluation and a comprehensive dataset that also includes ground truth information.

## II. Background

In this study, two prominent approaches for occupant estimation have been focused on. The relevant preliminaries for these two are highlighted.

### A. Sensor Properties

Certain sensors respond to the changes in environmental phenomena due to the presence of people. Other than the commonly used *motion sensors*, sensors like temperature, humidity and $CO_2$ also respond to human occupancy. Among them, the most significant is the $CO_2$ concentration in the air. Most of the studies [9], [10] basically modeled the $CO_2$ measurement in the indoor environment as,

$$C'_k = f(c_{k-l:k}, f^v(.)) \tag{1}$$

Here, $C'_k$ is the level of $CO_2$ measured during the instance $k$. $f(.)$ is a function on the parameters, $c_{k-l:k}$, the sequence of $CO_2$ measurements within the interval $[t_{k-l}, t_k]$ of length $l$ and $f^v(.)$. $f^v(.)$ is a function for the relevant environmental factors such as the *ventilation* system. It must also be noted here that the effect of occupant dynamics takes certain duration, which is not necessarily insignificant, to be reflected on the level of $CO_2$, temperature and humidity.

### B. Radio Signal Properties

The characteristics of radio waves as transmitted by wireless devices have been effectively used for occupancy analysis. The radio waves from a transmitter reach the receiver traversing different paths. The wave that is reflected and/or scattered exhibit different characteristics than the ones with the line of sight (LOS) traversal. This can be leveraged by Received Signal Strength Indicator (RSSI) [6], usually interpreted as the power of the received signal and expressed in dBm. The presence of obstacles effect the power of the traversing signal and thus can be utilised for occupancy estimation. In recent approaches, the Channel State Information (CSI) has been given significant importance for studying the occupancy patterns. The CSI can be modelled as,

$$R_x = HT_x + n \tag{2}$$

where, $R_x$ and $T_x$ are the receive and transmit vectors, $H$ is the channel matrix and $n$ is the noise vector. The CSI parameter expresses the combined effect of factors such as, scattering, fading, and power decay during the traversal. Another approach these days uses MIMO (Multiple Input Multiple Output) communication (Fig. 4) between the transmitter and the receiver. For such a case, with $i$ receiver antennas and $j$ transmitter antennas, Eq. (2) can be expressed as,

$$\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_i \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & \dots & H_{1j} \\ H_{21} & H_{22} & H_{23} & \dots & H_{2j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{i1} & H_{i2} & H_{i3} & \dots & H_{ij} \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \vdots \\ T_j \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_i \end{bmatrix}$$

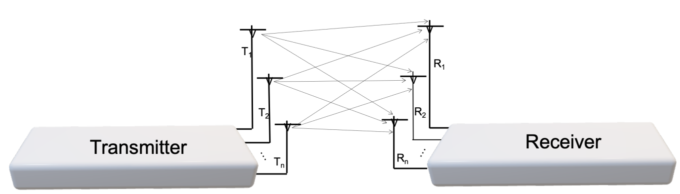



Fig. 4: MIMO communication scenario

The values, $H_{i,j}$ or in other words, the matrix, $H$, which usually has complex constituents, contain the channel state information and thus bear the effect of the human presence and/or movements. Again, a generic scenario that uses $K$ sub carrier frequencies, there will be a vector of $K$ different $H$'s at each instance and can be expressed as,

$$P = \{H^1, H^2, \dots, H^K\}$$

Such vectors can be formulated at pre-defined intervals thus generating a stream of $P$s.

$$Stream,\ S = \{P_1, P_2, \dots, P_t\}$$

## III. Literature Survey

The domain of occupancy detection and classification is a well explored area now. Camera based approaches [11] for studying human presence, usually have privacy concerns associated with it. Moreover, the installation and maintenance of required hardware also raises questions against its cost-effectiveness. Considering the ubiquitous sensing as the primary objective for this work, the camera based active sensing approaches are out of the scope and hence are not mentioned. This brings down the scope into two broad directions: sensor based approaches and radio frequency based approaches.

*1) Sensor based occupancy estimation:* Many sensors respond to the change in environmental phenomenon related to occupancy dynamics. Moreover, being cost-effective and pervasive in nature, the sensor based approaches have been studied extensively for quite some time. The motion sensors may generate a large no. of false alarms or even missed detection [12] due to various factors such as improper configuration, deployment location and other environmental effects. Various factors such as time of the day, controlled luminosity level, etc. effect the ambient lighting, and therefore, light sensor based approaches [12] have limitations associated as the occupancy and illumination may not be highly correlated. Other sensors for measuring temperature [13], humidity [14] and $CO_2$ [15] have been used for occupancy oriented studies. A multi-sensor based approach is expected to be more efficient and an inclination towards such a trend is quite visible. Many studies [16] have effectively utilised multi-sensor fusion rather than relying on a single type. Sensors such as sound sensors, motion sensors, vibration sensors, ultrasonic sensors, proximity sensors, passive infrared sensors, RFIDs etc. have also been used occasionally for occupancy analysis.

*2) Radio frequency based approaches:* A lot of interests can be seen in the literature to study wireless channel [17] for occupancy analysis. Earlier works mostly utilized the notion of counting connected devices [1]. But recent trends exploit the characteristics exhibited by the radio waves [18] due to the presence of obstacles or movement in the propagation path. It is pretty much observable that the effect on the radio wave is more visible if the obstacles are in motion and due to this, majority of the existing works focus on movements [19] of subjects. For being comparatively easier to collect, Received Signal Strength Information (RSSI) [4] of wireless signals have been preferred in many of the earlier works. But it is also observable that the increase in the no. of subjects sometimes results in uncertain behaviour of the energy change. The state of the art literature has growing interest towards the Channel State Information (CSI) as a potential indicator of the occupancy [19]. Challenge lies in the extraction of raw CSI data from a device. This usually requires specialised hardware and/or firmware installation and hence tampering an existing infrastructure.

*3) Measurement Goals:* There are *occupancy detection* approaches [20], which aim to detect presence or absence of individuals as binary prediction. Both sensor [20] and RF based [21] approaches have been used for occupancy detection. The other approaches try to address the problem in multi-class terms [22] and evaluate overall gathering (such as low moderate, high etc.). Another set of approaches try to estimate the exact count of the occupants. Such a problem can be addressed as both classification [23] as well as a regression problem [24]. In literature, another set of works aim to identify *groups* within a gathering of people. Some recent studies have explored identification of groups with respect to a defined *context*. For instance, collocation based social interactions has been explored in [25]. Performance evaluation in varying scenarios both indoors and outdoors has been reported in [26]. It is observable that in all of these, the context of the group formulation has been exploited by acoustic information. Whereas, WiFi signal data is used to exploit physical collocation [27]. However, at times capturing audio signals could be an invasion to the privacy. Moreover, intrusive factors such as installation of specialized software on users' devices might pose imposition challenges.

## IV. System Model and Assumptions

This section elaborates the system conceptualization. Whenever there is an *object of interest* (referred as the *objects* hereafter) in a room, it can trigger varied level of curiosity among *individuals* (referred as the *subjects* hereafter). Therefore, first we try to address, how to temporally relate a given subject to a specific object, in a scenario. Given a set of objects $N = \{n_1, n_2, \dots, n_k\}$, for each object, $n_i$, there exists a radius, $R_{n_i}$, within which the influence of the object $n_i$ is applicable. The value of $R_{n_i}$ may change with time due to the accumulation of more subjects. Therefore, $R_{n_i}$ at time $t$ can be stated as [2],

$$\begin{aligned} \Delta R^t_{n_i} &= \frac{-2R^I_{n_i} + 1 + \sqrt{(2R^t_{n_i} - 1)^2 + 4G^t/\pi}}{2} \\ R^t_{n_i} &= R^I_{n_i} + \mu_e \Delta R^t_{n_i} \end{aligned} \tag{3}$$

where, $R^I_{n_i}$ define the initial radius of the area within which the influence of $n_i$ exist. $\Delta R^t_{n_i}$ is the change in the radius of the area due to the accumulation of new subjects. $G^t$ is the number of gathered subjects at time $t$, $\mu_e$ is the influence coefficient of the gathering. Specifying a maximum value of $R^t_{n_i}$ as $R^{MAX}_{n_i}$ results in a zone marked by $Z_{n_i}$ encircling the object $n_i$ within which the subjects need to exist in order to be mapped to the object $n_i$ as shown in Fig. 5. $Z_{n_i}$ defines the coverage of sensory measurements corresponding to $n_i$ Next, we address, how the environment is affected by the

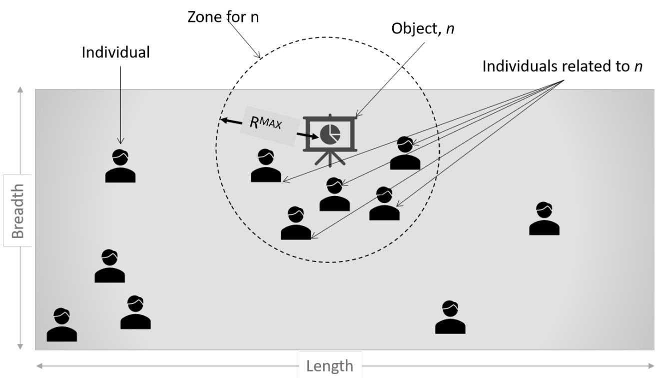



Fig. 5: Zoning

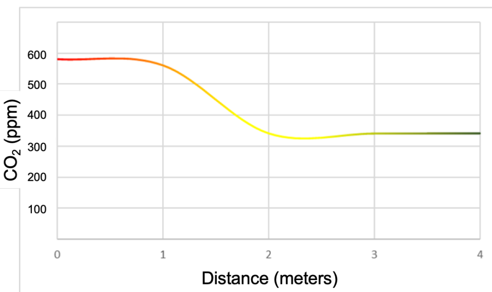



Fig. 6: The plot shows measured change of $CO_2$ concentration with increase in distance from the actual gathering (four measurements performed at an interval of 1 meter)

formulation of smaller gatherings. Each subject contributes to the concentration of $CO_2$ within the area designated by $Z_{n_i}$. Change in the $CO_2$ concentration is caused by the emission by the subjects. An observed state of $CO_2$ concentration due to the presence of 3 subjects within a close proximity of an object is shown in Fig. 6. Here, the measurement is performed once the 3 subjects remained in the particular location for at 3 least minutes straight while exhaling $CO_2$. Distance is the primary factor that isolates a group of subjects within the radius of influence of an object from the others. The level of $CO_2$ concentration decreases as we move further away from the object. This behaviour can be captured using a diffusion model for $CO_2$ concentration with the assumption that the concentration of environmental $CO_2$ is lower than that at the source of exhaling. We may consider a circle surrounding the source at which $CO_2$ is generated. Let $C$ be the concentration of $CO_2$ in the surrounding region of the source. Therefore, $C$ is a function of radial distance $r$ of the point in observation at time $t$ of the observation and thus can be expressed as, $C(r, t)$. Let $C_o$ be the value of $C$ at $t = 0$ i.e. the initial value. Let $C_s$ be the value of $C$ at the point where the source, $S$ is placed. Therefore, by the nature of diffusion of gases, $\frac{\delta C}{\delta r} \propto -C$ and $\frac{\delta C}{\delta t} \propto -C$. $\frac{\delta C}{\delta r}$ is the gradient of concentration while $\frac{\delta C}{\delta t}$ is the rate of transmission. Therefore, combining these two factors, we can write,

$$C = -k_1 \frac{\delta C}{\delta r} - k_2 \frac{\delta C}{\delta t},$$

where, $k_1$ and $k_2$ are two constants of proportionality. On solving these two partial differential equations we get,

$$C = C_1 \exp\left(\frac{-r}{k_1}\right) + C_2 \exp\left(\frac{-t}{k_2}\right) \tag{4}$$

Initially, at $t = 0$ we had, $C = C_o$ at $r = R$, signifying the surroundings and $C = C_s$ at $r = 0$ ($\because C_s > C_o$). With these conditions we get,

$$C_1 = \frac{C_s - C_o}{1 - \exp\left(\frac{-R}{k_1}\right)}, \qquad C_2 = \frac{C_o - C_s \exp\left(\frac{-R}{k_1}\right)}{1 - \exp\left(\frac{-R}{k_1}\right)}$$

Therefore, from Eq. 4, we can write,

$$C = \frac{C_s - C_o}{1 - \exp\left(\frac{-R}{k_1}\right)} \exp\left(\frac{-r}{k_1}\right) + \frac{C_o - C_s \exp\left(\frac{-R}{k_1}\right)}{1 - \exp\left(\frac{-R}{k_1}\right)} \exp\left(\frac{-t}{k_2}\right)$$

As more and more people gather, the no. of sources of $CO_2$ increase within the radius of influence of an object. thus, for, $n$ sources (subjects), the diffusion can be expressed as,

$$C = \frac{nC_s - C_o}{1 - \exp\left(-\frac{R}{k_1}\right)} \exp\left(-\frac{r}{k_1}\right) + \frac{C_o - nC_s \exp\left(-\frac{R}{k_1}\right)}{1 - \exp\left(-\frac{R}{k_1}\right)} \exp\left(-\frac{t}{k_2}\right) \tag{5}$$

provided, $0 \leq r \leq R$, $k_1 \geq 0$, $k_2 \geq 0$, $C_s > C_o$, $n > 0$ and $n \in N$. When, plotted on a 3-dimensional space, the expression results in the surface as shown in Fig. 7. This surface duly resembles the plot achieved by the actual measurements shown in Fig. 6. This model demonstrating a space-time continuum characterises an estimated area encircling an object within which the concentration of $CO_2$ is expected to *elevate* because of the gathering. Individual gatherings can have individual $CO_2$ elevation patterns. With formation of groups, such observed level of $CO_2$ can be different at different locations within the same room. Fig. 8 shows one such scenario with four points of interest where the level of $CO_2$ is observed. The state of the groups remained constant for at least 3 minutes prior to the measurements.

There are certain fundamental factors which also get involved as per our observation. For instance, due to the absence of uniform ventilation in most of the cases, the $CO_2$ tend to drift towards the corner areas of the room thus increasing the concentration eventually. Similar idea for diffusion applies to the modelling of parameters such as temperature and humidity as they also tend to diffuse with distance. For simplicity, the possible overlapping of the radii of influence of different objects has been ignored.

It must be noted that the change in the above mentioned factors due to its slow response can be effective when the subjects are stable relative to $Z_{n_i}$ for a relatively prolonged

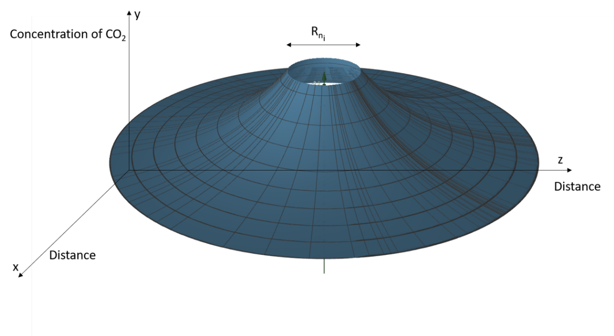



Fig. 7: Plot shows $CO_2$ level and its diffusion off a radius of influence, $R_{n_i}$ at time instance, $t = T$. The concentration level lowers gradually with the distance from $R_{n_i}$
.

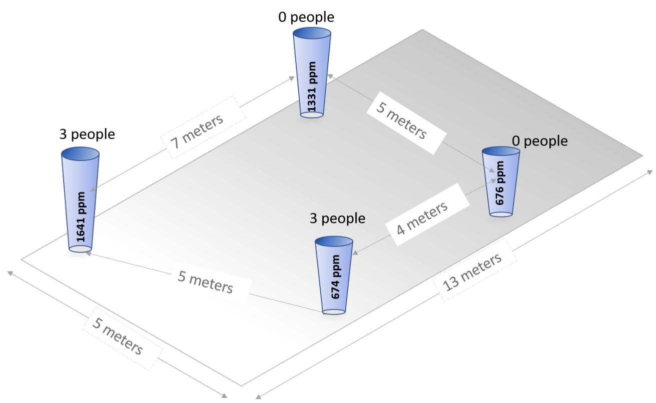
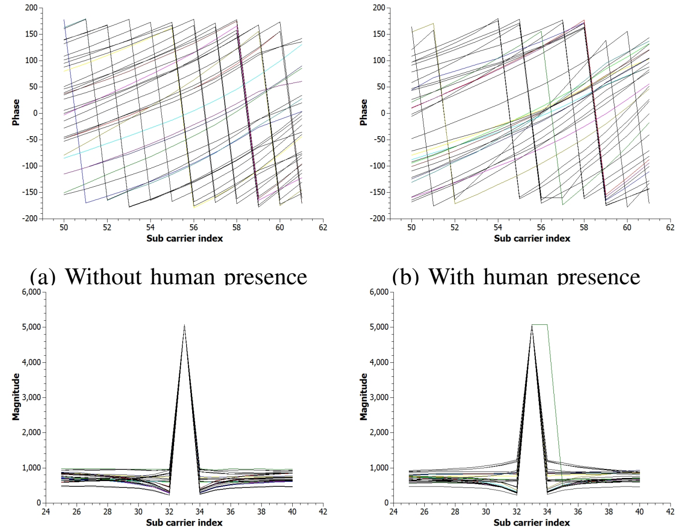



Fig. 8: Plot shows level of $CO_2$ at different points of interest at about 5 ft above ground level within a room with different groups, at some time instance $t = T$

duration. Moreover, other influencing factors such as presence of air conditioning system, un-uniform ventilation significantly affects these measurements. That said, apart from complementing the above, we also need to answer, how to detect and quantify the presence itself. To address this, we collect CSI information of the transmission waves effected by the presence of the subjects. The raw CSI measurements that we collect has 64 sub-carriers. For instance, once the phase and magnitude are extracted; the response to human presence is observable in phase for sub-carrier indices 50 to 60 (Fig. 9b) as the curves rise steeper with presence of people as compared to without people (Fig. 9a). Similarly effects on magnitude is observable in Fig. 9c and Fig. 9d

The presence of subjects and their dynamics effect the CSI parameters of the radio transmissions. There could be basically two situations how the CSI gets effected. One, when new subjects are introduced in the environment and they are stationary. As the received signal undergoes multi-path traversal, it is effected by various obstacles in its path as shown in Fig. 10. The frequency response can be expressed as [28],

$$H(f;t) = \sum_i a_i(t) e^{-j2\pi f \tau_i(t)}$$




(a) Without human presence (b) With human presence

(c) Without human presence (d) With human presence

Fig. 9: Example of impacted CSI parameters on a set of sub-carriers

Here, $a_i(t)$ is the amplitude attenuation factor, $\tau_i$ is the propagation delay and $f$ is the carrier frequency. The second situation is the effect caused by the movements of the obstacles and/or transmitters or receivers. The movement induces *Doppler's Effect* on the wave propagation. Thus, the frequency response for a sub carrier $k$ can be expressed by including the Doppler's shift as a function of time instance, $T$ [29],

$$H_k(T) = \sum_i a_i e^{-jw_k \frac{d_i}{c}} e^{jw_k \frac{v_i}{c}} T + \epsilon$$

where, $c$ is the propagation speed of the wave, $\epsilon$ is the measurement error and $w_k \frac{v_i}{c}$ is the Doppler shift. The amplitude and phase are affected by the obstructions and environment dynamics.

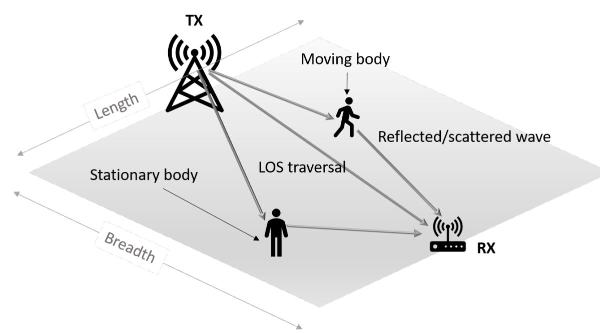



Fig. 10: Obstruction on multi-path traversal

## V. Experimental Setup and Implementation

To evaluate the performance of the proposed approach, we deploy a room-scale indoor testbed inside the naturally ventilated cafeteria of our institute's main building. We detail the setup below.

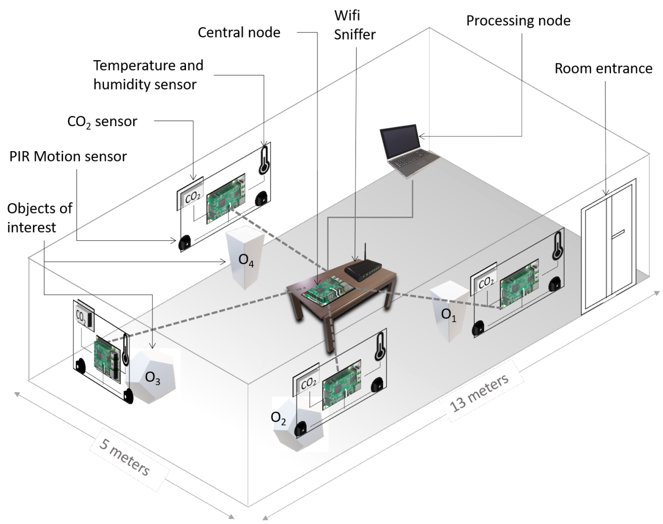



Fig. 11: Deployment of testbed

*1) Testbed Setup and Data Collection:* To address the curiosity and its influence, four different *objects* $O_1$, $O_2$, $O_3$, $O_4$ are placed at different locations of the room, as shown in Fig. 11. A wall-mounted *Sensing Box* is placed at close proximity of these objects about 5 ft above ground. Each Sensing Box has a Raspberry Pi Model 3B/B+, a set of HC-SR501 PIR motion sensors (aimed at different directions), a MH-z19 $CO_2$ sensor, a DHT22 Temperature and Humidity sensor. Considering the response time, sensitivity of these low cost sensor modules and to avoid heating due to rapid polling, the DHT22 sensors are configured to generate data at intervals of 3 seconds and the MH-z19s to generate at 60 seconds. The motion sensors were calibrated to sense the environment within the zone, $Z$ of $O_i$. Each Sensing Box transmits the sensory data collected by the connected sensory modules to a central node every 2 seconds using 802.11 based communication. The central node handles time-stamping and correlating the sensory measurements as well as maintaining time synchronization of the deployed Sensor Boxes. A sniffing device placed at the close proximity of the central node collects CSI information of the transmissions done by the Sensor Boxes but has no access to the payload contents. A passive sniffing device, as we introduce, does not require tempering with the existing devices already in communication and can be configured independently. Same wireless channel (channel 1 in our case) is used by all communicating devices. A sniffing device can monitor only one channel at a time. Therefore, we avoid dedicated channel configuration since, it will require more sniffing devices with increasing no. of sensor boxes. Moreover, the channel response patterns to human presence are different. The collected data is processed in an edge server formulated by a PC and connected to the central node. It is observable that the setup tries to maintain minimal disturbances to the subjects within the room or their movements. A volunteer is given task to collect the ground truth information at the site. The ground truth information has the no. of subjects within the zones defined by a radius of influence for *each object*.

Data is collected for sixty days which also include holidays (unless access to the room is restricted) during both office hours (9:00 to 17:00) as well as after hours (17:00 to 20:00) depending on the availability of people as well as time restrictions imposed by the institute. The room observes general gatherings throughout the day but the gathering is more during the breaks and after dusk (17:00 to 18:00) as more people hang out and have snacks during these times. As per our observation, a single zone observed up to six persons at any particular instance, whereas, the minimum number is zero.

*2) Feature Formulation:* The collected data is pre-processed before training as shown in Fig. 12. The CSI information with 64 sub-carriers is processed to extract the phase and magnitude vectors. Similarly, the sensor streams undergo removal of noisy measurements such as sudden spikes in measurements, etc. It must be noted that the data generation intervals of the sensors vary. Therefore, given an interval $[t, \delta t]$, if a sensor $s_1$ generates $n$ values, a sensor $s_2$ may not generate $n$ values. Therefore, in order to attain a specific granularity within an interval, *interpolation* is necessary to fill the missing measurements. The resultant is a time-series data at every second. A feature vector eventually consists of sensory measurements within a formulated window and the phase and magnitude vectors extracted from the CSI information.

*3) Machine Inference:* At any instance, 4 inference models are trained corresponding to each object. For training, a set of 4000 measurement instances were used. The trained models are then tested on generated stream of collected data at specific intervals as discussed in section V-1.

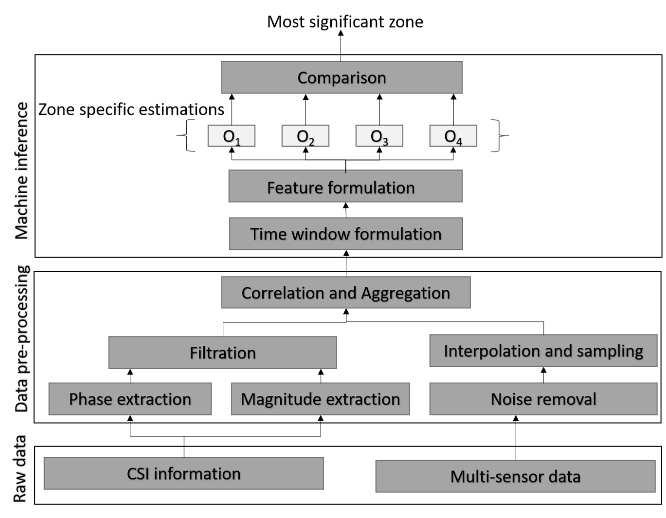



Fig. 12: The processing of collected data

## VI. Results and Discussion

To the best of our knowledge, no existing work specifically study and compare individual gathering sizes using passive sensing. Hence, we limit our experimental study towards performance analysis rather than a comparison. The crowd dynamics in our context is not something that change in rapid succession. Thus, a very short interval of estimation

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 86.67 | 92.0 | 85.34 | 62.67 |
| $O_2$ | 73.44 | 98.44 | 76.56 | 79.69 |
| $O_3$ | 88.73 | 97.18 | 95.77 | 74.65 |
| $O_4$ | 71.62 | 85.14 | 74.32 | 56.76 |

TABLE I: Window size 5 seconds

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 80.43 | 93.48 | 93.48 | 67.39 |
| $O_2$ | 56.25 | 90.63 | 84.38 | 78.125 |
| $O_3$ | 81.58 | 97.37 | 89.47 | 73.68 |
| $O_4$ | 76.74 | 81.40 | 81.40 | 67.44 |

TABLE II: Window size 10 seconds

| Model | Window: 5 secs | Window: 10 secs |
|---|---|---|
| LR | 53.13 | 43.75 |
| RFR | 93.75 | 81.25 |
| LR | 68.75 | 62.5 |
| PR | 59.38 | 59.38 |

TABLE III: Estimating largest gathering

Performance of the ML models while counting upto 3 individuals in a zone

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 22.22 | 61.8 | 31.94 | 13.88 |
| $O_2$ | 85.41 | 92.36 | 91.66 | 43.05 |
| $O_3$ | 23.33 | 60.0 | 73.33 | 25.55 |
| $O_4$ | 57.74 | 80.28 | 76.05 | 35.21 |

TABLE IV: Window size 5 seconds

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 19.45 | 66.67 | 33.34 | 18.06 |
| $O_2$ | 81.94 | 77.78 | 91.67 | 48.62 |
| $O_3$ | 34.72 | 61.11 | 70.83 | 18.05 |
| $O_4$ | 47.22 | 72.22 | 65.28 | 37.5 |

TABLE V: Window size 10 seconds

| Model | Window: 5 secs | Window: 10 secs |
|---|---|---|
| LR | 68.06 | 55.56 |
| RFR | 89.58 | 73.61 |
| LR | 68.75 | 56.94 |
| PR | 50.0 | 52.78 |

TABLE VI: Estimating largest gathering

Performance of the ML models while counting upto 6 individuals in a zone

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 88 | 92 | 85.33 | 66.66 |
| $O_2$ | 71.87 | 100 | 78.13 | 81.25 |
| $O_3$ | 88.73 | 97.18 | 95.77 | 76.05 |
| $O_4$ | 71.62 | 85.13 | 74.32 | 62.16 |

TABLE VII: Interval of 5 sec with window size of 5 sec

| Obj. | LR | RFR | LoR | PR |
|---|---|---|---|---|
| $O_1$ | 80.43 | 97.83 | 93.48 | 69.56 |
| $O_2$ | 37.5 | 93.75 | 75.0 | 68.75 |
| $O_3$ | 81.58 | 97.37 | 89.47 | 73.68 |
| $O_4$ | 76.74 | 81.39 | 81.39 | 67.44 |

TABLE VIII: Interval of 10 sec with window size of 10 sec

Performance of the ML algorithms while predicting state-change of a gathering

does not capture the favourable states of the environment effectively. Similarly, a very large interval of estimation may end up reflecting more than one event of gathering dynamics. Therefore, a trade-off must be made in order to select the measurement intervals. On careful observation, we conclude that the state of a gathering must persist for at least 5 seconds in order to be of significance. That means, if a new individual appears and disappears from a zone within a time-span of 5 seconds, he/she is just a mere passerby. This interval can be argued upon depending on use case scenarios. First we predict individual sizes of gatherings with respect to each object (Obj.). Four different Machine Learning (ML) models were studied: Linear Regression (LR), Random Forest Regression (RFR), Logistic Regression (LoR) and Polynomial Regression (PR) for predicting the size of gathering per zone. Two cases were studied. In the first case situation, each zone observed upto 3 people at any instance of time. TABLE I shows the accuracy in predicting the size of gatherings in individual zones measured at 5 seconds interval. And TABLE II shows the same with 10 seconds interval. Finally, accuracy in predicting the largest gathering is shown in TABLE III.

In the second case, the room environment observed upto 6 people in a single zone at any instance of time. The corresponding values are shown in TABLE IV, V, VI. And finally, prediction of crowd dynamics are shown in TABLE VII and TABLE VIII. The values show how accurately the models predict a changed or unchanged state of gatherings. The state of a gathering within a zone of an object change when one or more subjects either leave or appear within the zone. Again, we consider that a state change is of concern only if the changed state persists at least for 5 seconds.

One major issue, that affects the performance of the prediction is the fact that manual collection of ground truth cannot document instantaneous environment change. It must be noted that there also exist individuals within the environment who do not belong to any of the object specific zones and have some effect on the measurements. Moreover, the ground truth is collected by volunteers and thus, nominal human error in boundary cases while collecting are quite likely.

## VII. Conclusion

In this work, we have addressed a problem of passive and interest-oriented crowd estimation. As people gather in smaller groups driven by curiosity, we try to answer if it is possible to estimate the size of the groups by a contextually passive approach that offers minimal intrusion to existing infrastructure. The results of testbed experimentation show that the proposed method can offer an accuracy higher than 90% for crowd estimation for a given scenario.

## VIII. Acknowledgement

This work has been partially supported by SERB, DST, Government of India through project no. ECR/2017/000813.

## References

[1] Z. Wang, T. Hong, M. A. Piette, and M. Pritoni, "Inferring occupant counts from wi-fi data in buildings through machine learning," *Building and Environment*, vol. 158, pp. 281–294, 2019.

[2] Y. Li, K. Li, and L. Wang, "Crowd behavior simulation model based on information-gap theory," in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, IEEE, 2021, pp. 590–594.

[3] S.-C. Cheung, M. V. Venkatesh, J. K. Paruchuri, J. Zhao, and T. Nguyen, "Protecting and managing privacy information in video surveillance systems," in *Protecting Privacy in Video Surveillance*, Springer, 2009, pp. 11–33.

[4] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only wifi power measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1381–1393, 2015.

[5] H. Choi, T. Matsui, M. Fujimoto, and K. Yasumoto, "Simultaneous crowd counting and localization by wifi csi," in *International Conference on Distributed Computing and Networking 2021*, 2021, pp. 239–240.

[6] S. Depatla and Y. Mostofi, "Crowd counting through walls using wifi," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*, IEEE, 2018, pp. 1–10.

[7] A. Franco and F. Leccese, "Measurement of co2 concentration for occupancy estimation in educational buildings with energy efficiency purposes," *Journal of Building Engineering*, vol. 32, p. 101 714, 2020.

[8] V. Chidurala and X. Li, "Occupancy estimation using thermal imaging sensors and machine learning algorithms," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8627–8638, 2021.

[9] C. Basu, C. Koehler, K. Das, and A. K. Dey, "Perccs: Person-count from carbon dioxide using sparse non-negative matrix factorization," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 987–998.

[10] T. Ekwevugbe, N. Brown, V. Pakka, and D. Fan, "Real-time building occupancy sensing using neural-network based sensor network," in *2013 7th IEEE international conference on digital ecosystems and technologies (DEST)*, IEEE, 2013, pp. 114–119.

[11] S. Petersen, T. H. Pedersen, K. U. Nielsen, and M. D. Knudsen, "Establishing an image-based ground truth for validation of sensor data-based room occupancy detection," *Energy and Buildings*, vol. 130, pp. 787–793, 2016.

[12] G. Violatto, A. Pandharipande, S. Li, and L. Schenato, "Classification of occupancy sensor anomalies in connected indoor lighting systems," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7175–7182, 2019.

[13] D. N. Wagner, A. Mathur, and B. E. Boor, "Spatial seated occupancy detection in offices with a chair-based temperature sensor array," *Building and Environment*, vol. 187, p. 107 360, 2021.

[14] M. Naqi, S. Lee, H.-J. Kwon, *et al.*, "A fully integrated flexible heterogeneous temperature and humidity sensor-based occupancy detection device for smart office applications," *Advanced Materials Technologies*, vol. 4, no. 12, p. 1 900 619, 2019.

[15] M. Gruber, A. Trüschel, and J.-O. Dalenbäck, "Co2 sensors for occupancy estimations: Potential in building automation applications," *Energy and Buildings*, vol. 84, pp. 548–556, 2014.

[16] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," *Building and Environment*, vol. 180, p. 106 966, 2020.

[17] E. Longo, A. E. Redondi, and M. Cesana, "Accurate occupancy estimation with wifi and bluetooth/ble packet capture," *Computer Networks*, vol. 163, p. 106 876, 2019.

[18] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freecount: Device-free crowd counting with commodity wifi," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, IEEE, 2017, pp. 1–6.

[19] C. Tang, W. Li, S. Vishwakarma, K. Chetty, S. Julier, and K. Woodbridge, "Occupancy detection and people counting using wifi passive radar," in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, 2020, pp. 1–6.

[20] B. Abade, D. Perez Abreu, and M. Curado, "A non-intrusive approach for indoor occupancy detection in smart environments," *Sensors*, vol. 18, no. 11, p. 3953, 2018.

[21] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with svm and csi fingerprinting," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7990–7999, 2017.

[22] Z. Chen, R. Zhao, Q. Zhu, M. K. Masood, Y. C. Soh, and K. Mao, "Building occupancy estimation with environmental sensors via cdblstm," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 12, pp. 9549–9559, 2017.

[23] A. Sharma, J. Li, D. Mishra, G. Batista, and A. Seneviratne, "Passive wifi csi sensing based machine learning framework for covid-safe occupancy monitoring," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, 2021, pp. 1–6.

[24] Y. P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, and M. A. Little, "Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction," in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 2016, pp. 1016–1027.

[25] J. Baker and C. Efstratiou, "Next2me: Capturing social interactions through smartphone devices using wifi and audio signals," in *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2017, pp. 412–421.

[26] S. Das, S. Chatterjee, S. Chakraborty, and B. Mitra, "Groupsense: A lightweight framework for group identification," *IEEE Transactions on Mobile Computing*, vol. 18, no. 12, pp. 2856–2870, 2018.

[27] ——, "An unsupervised model for detecting passively encountering groups from wifi signals," in *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2018, pp. 1–7.

[28] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[29] R. F. Brena, E. Escudero, C. Vargas-Rosales, C. E. Galvan-Tejada, and D. Munoz, "Device-free crowd counting using multi-link wi-fi csi descriptors in doppler spectrum," *Electronics*, vol. 10, no. 3, p. 315, 2021.