1-2018

Smart monitoring via participatory BLE relaying

Meeralakshmi RADHAKRISHNAN
Singapore Management University, meeralakshm.2014@phdis.smu.edu.sg

Sougata SEN
Singapore Management University, sougata.sen.2012@phdis.smu.edu.sg

Archan MISRA
Singapore Management University, archanm@smu.edu.sg

Youngki LEE
Singapore Management University, YOUNGKILEE@smu.edu.sg

Rajesh Krishna BALAN
Singapore Management University, rajesh@smu.edu.sg

DOI: https://doi.org/10.1109/COMSNETS.2018.8328213

Smart Monitoring via Participatory BLE Relaying

Meera Radhakrishnan, Sougata Sen, Archan Misra, Youngki Lee, Rajesh Krishna Balan
Singapore Management University
{meeralakshm.2014, sougatas, archanm, youngkilee, rajesh}@smu.edu.sg

***Abstract*—We espouse the vision of a smart object/campus architecture where sensors attached to smart objects use BLE as communication interface, and where smartphones act as opportunistic relays to transfer the data. We explore the feasibility of the vision with real-world Wi-Fi based location traces from our university campus. Our feasibility studies establish that redundancy exists in user movement within the indoor spaces, and that this redundancy can be exploited for collecting sensor data in an opportunistic, yet fair manner. We develop a couple of alternative heuristics that address the BLE energy asymmetry challenge by intelligently duty-cycling the scanning actions of individual devices. We evaluate the efficacy and tradeoffs of the proposed approaches by simulation experiments with real-world location traces.**

I. Introduction

There is considerable excitement around the vision of a smart city/campus, where sensors attached to objects such as garbage bins, vending machines and cafeteria seats, provide continuous reports on their status, such as the level of utilization of a garbage bin or the occupancy of a seat. Advances in miniaturization allow them to be deployed in almost any environment, in diverse form factors and on different objects.

Our deployment scenarios envision a very dense deployment of such low-energy, short-range BLE-equipped (Bluetooth Low Energy [1]) sensors attached to hundreds or thousands of commonly used objects, across an entire campus or city. Based on the application requirement, these sensors should have the capability of transferring back their sensed data frequently. However, transferring data from the sensors to a central IT platform remains an unsolved challenge: Wi-Fi is simply too energy intensive, fiber or Ethernet cabling does not work with potentially-movable objects (e.g., cafeteria seats or garbage bins), low-energy protocols such as LoRa have low bandwidth, and multi-hop wireless approaches such as 802.15.4 are difficult to maintain over extended deployments.

To overcome this limitation, we propose a human-centric approach, which we call *SmartABLE*[1], where smartphones carried by users are used to interrogate and collect data from *nearby* sensors in an *opportunistic* and *participatory* manner. The phones then transfer this data to the central IT platform over a conventional Wi-Fi or 4G interface. This approach eliminates the twin challenges of (i) networking complexity and (ii) high energy overhead by using phones (referred to as *mules*) as opportunistic one-hop relays that provide backhaul connectivity. The pervasive sensors just support simple, one-hop short-range communication.

The use of mobile data mules has been extensively studied in two main domains: (a) in wireless sensor networks (WSNs), where a mobile node travels through the sensor field acting as a collector [2], [3], and (b) in delay-tolerant networks (DTNs), where mobile nodes carry a packet towards its destination [4], [5]. Our envisioned scenarios differ in the following ways:

- Unlike WSNs, which focus on directing a mule's path to optimize some collection metric, *SmartABLE* is opportunistic, where the data collection is piggybacked on the non-deterministic movement of hundreds of users. Hence, it is unclear if the natural ebb and flow of human movement will provide sufficient coverage throughout the day and across all locations on a campus. However, *SmartABLE* is also *centrally coordinated*, with the scanning schedule of mules being intelligently orchestrated by the backend. Also, to work in predominantly indoor environments, this coordination must work with the coarse-grained location accuracy (typically 8-10 meters) typical of currently deployed indoor localization systems.
- Smart city/campus monitoring scenarios usually require tighter guarantees on the *reporting frequency* than DTNs generally provide – for example, the campus facility manager may desire to know the status of garbage bins once every 15 minutes and the occupancy status of seats in the cafeteria every 5 minutes. This can be a challenge not just because of the time-varying movement patterns of individuals, but also due to possible fluctuations in the participation rate of users.

In this paper we analyze the feasibility of realizing a smart environment, where data mules can assist in relaying information from smart sensors to the backend. Additionally, we discuss about possibility of heuristic based energy reduction without compromising on the coverage. Through the analysis, we make the following **key contributions**:

- *Establish both the promise and problem of participatory data muling:* Using real-world indoor location traces of thousands of users in our *University campus* [6], we show that data mules are available at most publicly accessible parts of the campus, even at different times of the day. However, we also show that straightforward data muling scheduling solutions are problematic: having mules perform continuous BLE scanning consumes too much energy, while static load-sharing across different mules results in unacceptably long reporting gaps, especially from less-traveled locations or during off-peak hours.
- *Propose new scheduling heuristics:* We explore a family of heuristics for *SmartABLE* that schedule BLE scanning by a group of participating mules (user smartphones), which achieve different tradeoffs between the goals of high sensor coverage and low BLE-scanning energy cost.

[1]**Smart A**daptive **BLE** Relaying

These include: (a) a *myopic* approach that uses only *instantaneous* location of each available user to decide the scan schedule; (b) a *greedy* approach that uses coarse-grained predictions of a mule's future movement behavior (derived from location trace histories) to schedule users, and (c) an *energy-conscious* variant, of the greedy approach, that improves the fairness of the scanning load.

- *Evaluation of tradeoffs of proposed heuristics:* We evaluate the tradeoffs of the scheduling heuristics using simulations on real-world location traces of thousands of occupants at the *University campus*. We identify techniques which can cater to diverse application requirements, such as (a) providing frequent updates (once every 20 minutes) for at least 90% of the resources or (b) reducing the number of scanning phones by 15x as compared to approaches that provide frequent updates.

II. Related Work

Works that align closely with our vision of data muling in a smart campus lie principally in the areas of (a) data muling for sensor networks, and (b) human mobility-aware urban sensing.

Data Mules in Sensor Networks: The concept of Data Mules for collecting sensor data in sparse sensor networks was first introduced in [4]. The key idea was to utilize the motion of existing entities in the environment to collect sensor data. Whenever the entities were near a base station, they would transfer the data, thus making it suitable for delay tolerant scenarios. There have been many other works (e.g. [2], [3], [7], [8]) which used the data muling concept to relay data from WSNs deployed in various domains. However, unlike WSNs, we envision an opportunistic data collection approach that relays information from several IoT sensors, piggybacked on the non-deterministic movement of hundreds of users.

The use of human-carried mobile phones as data mules for sparse sensornet deployments was first explored in [5],where the authors have shown that both intentional and opportunistic mobility can be used for data muling. However, each mule effectively operated independently, with a pre-specified scanning interval, and the system had no pre-specified reporting frequency requirements for an individual sensor. Unlike [5], we specifically investigate how the scanning behavior of multiple mules can be *collectively* adapted, while taking into account explicit reporting frequency requirements. More recently, works such as [9], [10] have used mobile phones as data mules to opportunistically collect sensor data. In contrast, we specifically investigate how this approach can take advantage of the predicted movement of individual mules.

Human Mobility based Urban Sensing: Mobility patterns of humans (both indoor and outdoor) using cell tower information or GPS traces has been well-studied [11], [12], [13]. Additionally, [11] studied spatial variations in both daily human movements and the interactions/communication behaviors between individuals. However, they do not conduct studies to understand the temporal variations in movement and occupancy within indoor environments. Similar to our study, a coarse grained (both spatial and temporal) campus wide pedestrian mobility study was conducted [14]. However, the authors did not evaluate the spatio-temporal coverage at different locations, different times and with different occupancy levels.

The feasibility of leveraging human mobility patterns for crowd-sourced urban sensing scenarios has been studied in some recent works [15], [16], [17]. In [15], the authors proposed a campus-scale crowd tasking platform that recommended simple tasks to users based on their predicted movement trajectories. [16] assumes that the user's path between source and destination is a free variable that we can modify to maximize coverage. This work also does not try to balance the energy utilization among users.

III. *SmartABLE*: Enabling Smart Campus Scenarios

Figure 1 illustrates our vision for a smart campus where sensors are attached to a variety of campus resources and objects. For example, accelerometers attached to doors (bathroom, study rooms, etc.) count how many times a door has been used, capacitive sensors under tables indicate whether a particular seat is occupied, and light sensors indicate whether lights in rooms have been left on. These sensors incorporate BLE stacks that broadcast beacons (advertisements) with embedded sensor information. Note: these sensors are now commercially available – e.g., Estimote^{TM} beacons broadcast advertisement beacons with embedded raw accelerometer and temperature sensor readings, or contextual data (e.g., motion status).

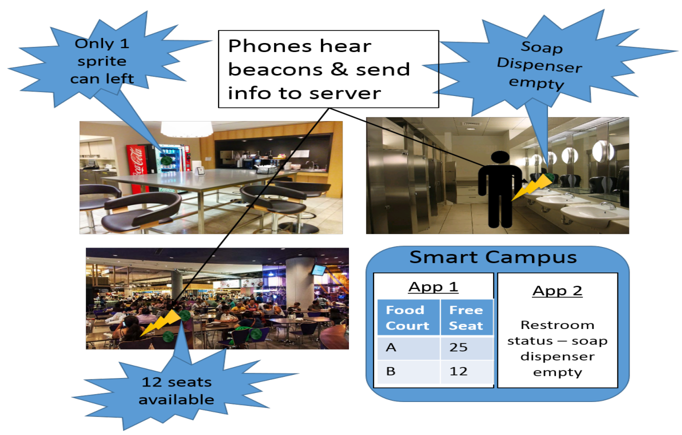


Fig. 1: Smart Campus Scenario

Individuals carrying their personal mobile devices move about the campus freely, based on their daily lifestyle patterns. These mobile devices have their Bluetooth interfaces activated, performing BLE scans either continually or intermittently. Whenever a scanning mobile device comes near such a BLE-equipped sensor, it picks up the sensor data or context about the associated object from the corresponding BLE advertisements. The mobile device then transmits this information back to a backend infrastructure, using its conventional wireless network interfaces (e.g., using Wi-Fi or 4G). At the backend, the data collected is analyzed and aggregated into appropriate portals, and also published to relevant applications which have subscribed for the information. For example, a restroom monitoring application will utilize the continual set of updates about the motion state of a specific restroom stall door, and then alert the cleaning staff when the number of “door open” events exceeds a specified threshold.

The mobile devices thus act as mobile relays, collecting sensor data about objects that happen to lie on their paths, and then transferring such data to a backend infrastructure. Of course, the periodicity/frequency with which the backend server receives reports from an individual sensor will depend on the overall occupancy pattern of the campus: the larger the gap between successive individuals transiting past a specific object, the greater the gap between successive updates. Our hypothesis is that this approach would, however, be very effective for campuses and commercial spaces that have high occupancy density (as is the case in many urban environments in Asia). Moreover, this model, of using mobile devices as a relay for transferring data from a set of pervasive sensors, can apply to a variety of other indoor and outdoor environments: for example, in a hospital, where the mobile devices of hospital staff are used to continuously collect reports on the status and location of hospital equipment equipped with such BLE-enabled tags.

A. Design Goals

In view of the application context, the proposed *SmartABLE* framework should meet the following design goals:

(1) Frequent Status Reports: The primary goal of *SmartABLE* is to ensure that in the campus, the server frequently receives sensor data from every deployed sensor. The frequency at which data is received at the server will be application and sensor specific and will vary based on time of day.

(2) Energy Conservation: Since an individual’s smartphone has to perform the BLE scan and relay any sensor data to the server, *SmartABLE* should attempt to minimize the energy consumption of an individual’s smartphone.

(3) Fairness: As there are multiple individual smartphones that will be scanning the campus to ensure frequent generation of sensor reports, *SmartABLE* has to ensure that the battery level depletion across the smartphones is as fair as possible. However in certain scenarios, where only one smartphone can update the status of a sensor (i.e., only one smartphone is near the sensor), it may be OK to impose short-term unfairness, unless the phone’s battery has dropped below a critical threshold.

B. Potential Challenges

For *SmartABLE* to be an effective approach for smart campus monitoring, it is essential that this model of opportunistic mobile relaying be capable of providing some sort of guarantees about the reporting frequency. Clearly, *SmartABLE* can offer only soft real-time guarantees, as its performance is dependent on the pattern of on-campus movement of all participating individuals. For a wide variety of practical applications, such soft guarantees are likely to be sufficient–e.g., for restroom monitoring, it should be OK to receive reports on individual restroom stalls only approx. 15-20 minutes or so.

Given a set of participating individuals, the best performance (in terms of frequency of reports) would be obtained when each mobile device performed its BLE scan continuously (at all times), as this would result in the largest number of readings obtained from all objects. However, this approach is not feasible–due to the asymmetric energy expenditure of BLE scanning [18], individual smartphones would experience unacceptably high battery drainage. Two key insights suggest that we can avoid continuous scanning simultaneously by all participating devices: (i) high occupancy density provides redundancy, and should thus allow the scanning load to be distributed among multiple collocated devices, and (ii) there is no extra benefit from overly frequent updates, as many smart campus monitoring applications require such updates only periodically.

C. *The* SmartABLE *Framework & Architecture*

These insights motivate our proposed *SmartABLE* framework. Figure 2 shows the functional architecture of *SmartABLE*. *SmartABLE* assumes the existence of an underlying location tracking infrastructure, that passively captures the movement history of all mobile devices on a campus. Such a location tracking solution is in fact an operational reality on our University campus, where a server-side Wi-Fi fingerprinting based indoor localization service has been deployed for the past 3 years [6], providing near-real time tracking of tens of thousands of Wi-Fi-enabled devices with an average location error of $\pm 6-8$ meters.

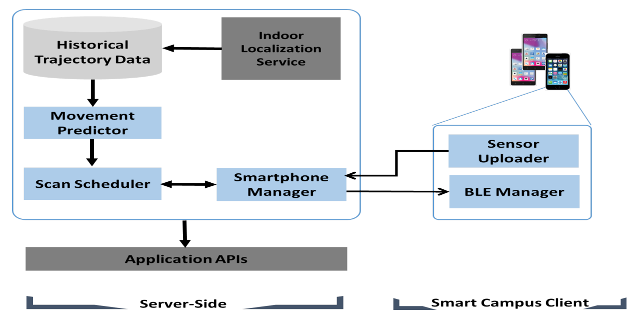


Fig. 2: *SmartABLE* Architecture for energy-efficient BLE scanning

The *Movement Predictor* component at the backend leverages upon such movement history to provide short-term predictions of the movement paths of each participating mobile device, and thereby derives the redundancy due to the likely co-location of multiple mobile devices. The *Scanning Scheduler* computes a set of *activation schedules* for such individual devices, while the *Smartphone Manager* component instructs each individual device to performing BLE scanning according to its specified schedule. The computed schedules: (a) provide a high likelihood that the interval between successive reports obtained about a campus object will not exceed a specified maximum threshold, and (b) distribute the energy overhead of BLE scanning across the set of participating devices, so that an individual mobile device does not incur an unacceptably high battery drain. Each participating mobile client includes a “SmartCampus client” that receives updated scanning schedules from the server-side “Smartphone Manager”. This client then performs the actual scanning using a “BLE Manager” component; the collected scan data is send back to the server using a “Sensor Uploader” component.

D. Research Questions

To systematically explore the feasibility of *SmartABLE* framework, we shall investigate the following questions:

- Given real-world movement patterns, how much redundancy is there in a densely occupied campus, and how frequently can we expect to receive updates about specific campus objects? More importantly, how much skewness/ variance of such coverage/redundancy is there at various times of the day?
- If such redundancy exists, how can one amortize the cost of relaying among different phones? More specifically, is it enough to duty-cycle individual devices according to a static schedule or to just utilize a fixed subset of devices (chosen randomly) in the scanning/relaying process or is any better strategy needed?
- If a smart *adaptive* strategy is needed, what is it? What are the tradeoffs between a myopic approach that uses only the current (instantaneous) device locations vs. an approach that utilizes short-term, but uncertain, predictions of device movement? What is the resulting energy cost vs. coverage gain (in terms of the distribution of the reporting frequency) that can be obtained with various scheduling heuristics?

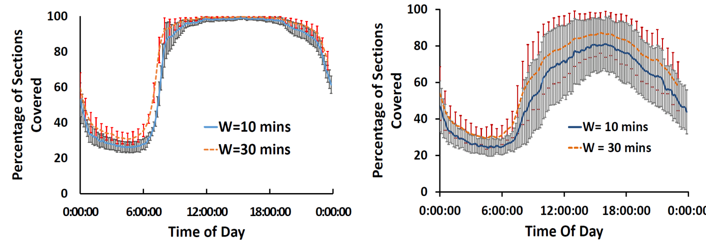
IV. Spatiotemporal Characteristics of User Movements

We first investigate the feasibility of our envisioned framework by conducting extensive studies to understand the spatio-temporal characteristics of users in the *University campus* from their indoor mobility data traces.

A. Dataset

To understand the movement behavior patterns in indoor environments, we utilize longitudinal traces of movement data obtained via our LiveLabs location service [6], which obtains the location of *any* Wi-Fi-enabled device by utilizing measured RSSI values on the uplink captured by multiple Wi-Fi APs, and thus does not require the installation of any application on individual mobile devices. The dataset contains movement traces of any Wi-Fi device which has been heard by an access point in our university campus. The indoor location service in our campus spans across five separate academic buildings, in addition to a connecting public concourses. We divided the entire area of the campus into *247* logical sections, where a section can be as small as 18 sq. m to as large as 108 sq. m. For our feasibility study, we compute the coverage (explained in Section IV-B1) based on these sections. Each section comprises multiple landmarks, with the inter-landmark distance being 3 to 6 meters. There are in total *2016 landmarks* across all the sections.

For the feasibility studies, we utilize Wi-Fi data captured over the period of March 2016. The real-time location system provides a location update for all available Wi-Fi devices once every 5 seconds. We consider the location traces from only devices that are heard at least t times during the month (we empirically set $t = 60$, implying a minimum observation duration of 5 mins over the entire month) and also are not stationary (stays in one section - e.g. laptops inside labs). We



(a) Coverage on Weekdays (b) Coverage on Weekends

Fig. 3: Temporal Variations of Coverage in University Setting

also considered only devices whose daily movement, averaged over the month, exceeded 3 sections. After such filtering, we had 16,907 unique devices for March 2016 that were considered as *regular* and *mobile* users.

B. Feasibility Study Metrics

For the study of real-world user movement behavior, we user two key metrics: (1) *coverage* and (2) *count*.

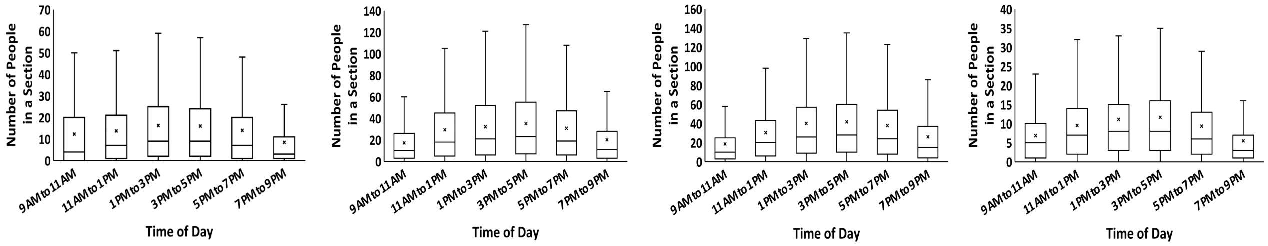
1) Coverage: Given a data time window (w), the *coverage* (C) is defined as the number of *covered* sections (n) divided by the total number of sections (N). A section $s \in S$ is considered *covered* if there exist a user $u \in U$ located in the section during a reporting period, w. $C = 1$ would indicate that the entire section can be monitored at least once by a user during the reporting period. Unless specified, we use $w = 600$ sec. as the default value for w. We explore how the coverage changes under various conditions (e.g., time of the day, data collection window, participation rate, etc.).

2) Count: We also compute the *count* of a section (Ct_s), i.e., the number of people found to be collocated in a specific section 's' within a single time slot 'w' (thus, $Ct_s = 0 \rightarrow$ s is uncovered). Count of a section will assist in understanding if there is enough user redundancy in each section at different times of the day. In case there is redundancy, *fairness* in terms of the data mules engaged in data collection can be achieved.

C. Results and Observations

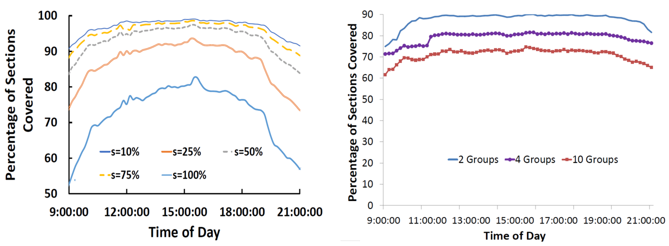
1) Mean Coverage and Effect of Window Size: We first study (a) the skewness/ variance of coverage/redundancy at various times of the day and (b) if the variation is similar for different location types.

Figures 3 shows the temporal variation of coverage on weekdays and weekends. From the figures, we can see that the coverage varies based on time of day and day of week. We find that the average coverage is greater than 97% between 11:30 AM and 7 PM in weekdays, i.e. most sections has either one or more people. The coverage drops during non working hours and is the lowest during the early morning, exhibiting a sharp rise from 7:00 AM. In case of weekends, there is no sharp rise in coverage. The coverage increases gradually and peaks around 3 PM. We also see that varying the window size does not have significant effect on coverage. When the window size is increased 3x (implying phones will have to scan 3x more), the maximum improvement in coverage is about 5%. Since the coverage during working hours of weekdays is similar for $w = 10$ minutes and $w = 30$ minutes, the proposed application



(a) Classrooms (b) Common Seating Areas (c) Open spaces (d) Office Spaces

Fig. 4: Coverage across different types of Locations in University Campus



(a) Random Scheduling (b) Round Robin Scheduling

Fig. 5: Coverage with subset of users (based on *random* and *round robin* scheduling) on weekdays in the *University campus*

scenarios could use a window size $w = 10$ minutes, so as to lower the overall energy consumption.

2) User Redundancy: To understand how many users are present in a section at a particular time and in turn obtain the redundancy, we divided the entire month's data into time slots of $w = 10$ minutes. We divided the university campus into four logical groups - (i) classrooms/seminar halls; (ii) common seating areas (iii) open spaces/corridors and (iv) office spaces. Figure 4 show the box plots for each of the logical locations for different times of the day. From Figure 4, we can see that in the campus, common areas have a median count of ≈ 20 users, while classrooms and office spaces have a median of ≈ 10 users indicating high redundancy. These results indicate that there exists enough user redundancy at different locations across the camous environment and is thus, promising for our envisioned scenarios and to ensure *fairness* across smartphones' battery depletion level.

3) Effect of Sub-sampling: We also explore how the coverage/count value changes when we drop our assumption of 100% participation of all users. Instead, we investigate two alternative naïve strategies that reflect reduced participation rates at any instant:

- *Random sampling:* In this approach, we randomly selected a specified percentage of the total pool of participants; the experiment was repeated 15 different times (with different randomized selections). For our analysis, we varied this participation percentage between $\{10\%, 25\%, 50\%, 75\%\}$.
- *Round Robin:* In this approach, we did not reduce the pool of participants but instead mimicked a time-slotted scheme, where the user were divided into N separate groups. BLE scanning was performed by all the members of a specific group in a particular slot, with the different groups taking turns across the slots. We experimented with $N = \{2, 4, 10\}$.

Figure 5(a) shows the variation for *random sampling* of users in the campus. From the figures we can see that when we reduce the sampling rate to 50%, the mean coverage drop is limited to 92%. This indicates that there is sufficient redundancy in user locations, showing the feasibility of using participatory sensing approach. On the other hand, when the sampling rate is much lower, i.e., 10 %, the coverage dropped to less than 70%. Of course, if the participant selection is skewed (e.g., greater proportion of recruits from a particular school), the coverage rate is likely to be lower. Similarly, in the case of time-slotted round robin scheduling (see Figure 5(b)), the coverage drops significantly when the users are divided into 10 buckets and each user bucket scans periodically in a round robin manner. This implies that naive random sampling and round robin sampling approaches would suffer when the participation rate is low, highlighting the needsfor an intelligent scheduling algorithm.

D. Summary of Key Insights

We summarize below the key insights from our study:

- The maximum coverage at our University Campus is high (at working hours across all weekdays), showing the feasibility of participatory BLE relaying.
- Redundancy exists in the user movement at any point in time in the campus and exploiting this redundancy will help in avoiding individual users from performing continuous scanning while still providing maximum/close-to-maximum coverage.
- Naïve subsampling methods such as random or round robin lead to significant drop in coverage. This requires us to come up with advanced strategies in selecting the users for data scanning and reporting.

V. Heuristics for Smart Scheduling

In Section IV, we show that the coverage varies across the day. All the studies in the previous section were done using a *Naïve* approach, where either all phones are always scanning or else are selected randomly at each scanning period. However this might not be the most efficient approach as scanning continuously is highly energy intensive. To minimize the energy consumption on phones, without compromising significantly on the coverage, we now propose a couple of heuristics for determining an improved scanning schedule.

A. Instantaneous Scheduling Approach

A straightforward approach of determining the set of scanning users is to use a *user's real time location* (obtained from real time location data) and have them scan for that particular location for the given time window. The available users, across all locations, at the *beginning* of every time slot are chosen, and we assume that the user will stay at his current location throughout the time window w. Thus, unlike the Naïve approach, this appoach does not consider the changing locations of a user when deriving the scanning schedule. Our previous studies [18] have shown that requiring each smartphone to continuously scan for BLE devices is infeasible as it imposes a very high energy cost. Hence, once a phone is selected as a mule at the beginning of the time slot, it keeps itself on with a fixed scan interval (T_s) and duty cycle (d_c). Considering optimized parameters identified in [18], for our studies, we fix (T_s) to *10 seconds* and (d_c) to *50%*. (Note: the parameter values can be varied depending on the application specific mobility characteristics).

We expect that the *Instantaneous* approach will be highly efficient in terms of the coverage. However, as this approach does not use any historical information to know a user's expected location at any given time, this approach will result in more people being scheduled to scan. Also, there are chances that (i) the coverage would suffer at low participation rates of users, (ii) users are not scheduled fairly (i.e., certain users being chosen for scanning multiple times over a day, while some other users not selected at all).

B. Greedy Approach

We next propose a simple greedy approach that utilizes *probabilistic predictions of user movement* in campus. The *Movement Predictor* component of *SmartABLE* framework predict a person's stay points for the next w minutes (w = 10 minutes) at *section level*, based on historical traces of individual user movement obtained from the server-side Wi-Fi based indoor location system functional at our campus.

We assume that the BLE beacons (sensors) are at *landmark* level (i.e., each *section* has multiple landmarks). Our location tracking is at a higher granularity of *section* level (as landmark level location predictions will be more erroneous). Thus, if a person p is predicted to visit location (section) l with a probability P_{pl}, we assume that probability of reading data from sensor i ($i \in l$) is P_s. Then, probability of person reading the sensor i is $P_{pl} * P_s$. Based on our empirical studies [18], considering the beacon miss rate, we set P_s to 0.8.

Utilizing the probabilistic predictions of user movement in campus, the proposed *Greedy* approach uses the *minimum dominating set* of users at any given time slot to cover the different locations. For this, the greedy approach first choose the phones which can cover the maximum number of sections at any point in time. Hence, other phones who are at similar locations during the same time could be discarded. The minimum dominating set approach is repeated until each sensor is covered with a probability $\geq P_c$ (threshold empirically set to 0.7). Also, as mentioned earlier, the selected phone will scan throughout the time window with the fixed scan interval ($T_s = 10 seconds$) and duty cycle ($d_c = 50\%$).

The greedy approach avoids scheduling redundant users– i.e., those who are at the same locations as that covered by the selected users. Thus, we expect that, compared to the *Naïve* approach (everyone available throughout the time window scans) and the *Instantaneous* approach, the *Greedy* algorithm could reduce the number of scanning users, thus minimizing the overall energy consumption.

C. Energy-Aware Greedy Approach

We also tested a simple variation of the *Greedy* approach, where a user would not scan in consecutive slots. This approach could reduce the battery drop of individual's smartphone. In this approach, a user can be selected for inclusion in the scan schedule only if he has not been scheduled in the last n_t time slots (we set n_t to 3 in our case). A weight/goodness-metric is assigned to the user based on the number of locations he scans: higher the number of locations scanned, higher is the weight. However, the weight is reduced depending on when the user last scanned. In this approach, if there is no other person to scan other than the person who scanned in the last time slot, then that person will be chosen automatically, thus ensuring coverage for the sensors in that section.

VI. Experimental Evaluation

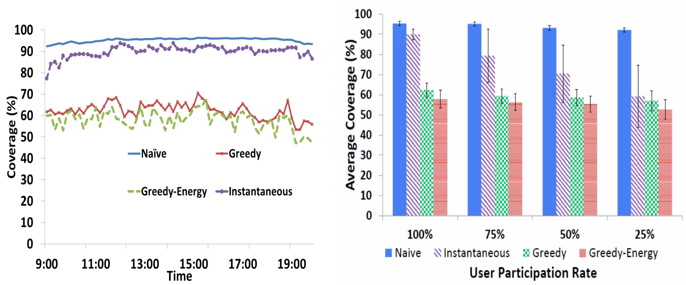
In this section, we present the evaluation and experimental results for each approach using *trace-based* simulation with real location traces. The scan schedules are obtained based on movement data predictions trained on historical location traces.

A. Trace-based Simulation

We evaluate the efficacy of different scheduling approaches using simulation experiments conducted with real location traces. In the experiments, the scan schedule for the greedy approaches is generated from 20,330 repeatedly visiting individuals, based on data during the working hours (9AM-8PM) on weekdays in February 2017. We divided the data into 10 minutes scan schedules (66 scans scheduled per day). The impact of the prediction-based schedules was tested using the movement data for the first two weeks of March 2017. As the sections vary in size, for the trace-based studies we assume that a sensor is associated with every *landmark* and compute the coverage and other evaluation metrics based on these landmarks.

To accommodate the uncertainty in the indoor location prediction approach, we use a 2 level system - where the currently predicted landmark by indoor localization will have a probability of 0.5, and the remaining probability of is distributed among the landmarks that are 1-hop away (i.e., each of the n 1-hop landmarks has a residency probability of $0.5/n$). In the trace-based studies, we indicate that a "beacon is read" only if the sum of the probabilities (across all [users, location data points]) exceeds a threshold P_c set to 0.7 (as mentioned earlier in Section V-B).

Based on our past studies [18], we make a simplifying simulation assumption that the probability of reading a sensor s is either 1 or 0, i.e., a beacon is either within or beyond reading



(a) Coverage at different times of a day

(b) Coverage at variable user participation

Fig. 6: Comparison of Coverage obtained by different scheduling approaches (a) across different time slots in a day and (b) at variables rates {100%, 75%, 50%, 25%} of user participation

	Naïve	Instantaneous	Greedy	Greedy-Energy
Average	6441	1967	122	114
Std. Dev	1478	489	18	15

TABLE I: Average number of users scanning in a day across all 10 minute time slots between 9am and 8pm

range. (Because of the high beacon reporting frequency, the probability of reading *at least one* beacon advt. is high, even if the miss rate is moderately high.)

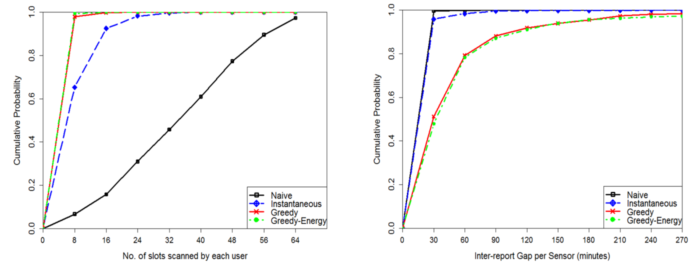
B. Evaluation Metrics

Based on the key design requirements for the *SmartABLE* framework, we use following evaluation metrics:

1) **Coverage:** The percentage of sensors covered out of total (one beacon at each of 2016 landmarks) in each of the 10 minute time slots over a day.
2) **Fairness:** The number of times each phone performed BLE scanning in a given day.
3) **Energy-efficiency:** The overall-energy consumed by *SmartABLE* in a day is computed by multiplying the total number of people scheduled for scanning in a day with the number of times they were scheduled to scan.
4) **Inter-report gap per sensor:** The gap in time (in minutes) in which a sensor report is heard from each sensor in a day.

C. Results

1) Coverage: Figure 6(a) show the comparison of coverage obtained by each of the approaches at different time slots across working hours of weekdays. From the figure, we observe that the *Instantaneous* approach is able to obtain a high coverage (as expected) of about 89% compared to the *Naïve* approach (without any sub-sampling) that achieves 95% average coverage. The *Greedy* scheduling approaches obtain a coverage of only about 62%. The reduction in coverage is mainly because of two reasons: (i) only the users currently present in campus and also in the scanning schedule (generated from training data) are considered, and (ii) certain scheduled users did not perform scanning as their BLE was manually turned off at the designated time (note again that the selected mules scan with a $T_s = 10 seconds$ and $d_c = 50\%$). However, the number of phones scanning is significantly fewer in the *Greedy* approach compared to the *Instantaneous* approach. Table I



(a) Fairness of users scanning

(b) Inter-report gap per sensor

Fig. 7: Comparison of different approaches for (a) *Fairness* and (b) *Inter-report gap*.

shows the number of phones scanning across 10 minutes time slots (in a day from 9 am to 8 pm). The *Instantaneous* approach uses more than 15x the number of phones, as compared to *Greedy*. These results also suggest that, if a higher coverage rate is desired, *Greedy* can simply be modified, to include a larger user pool, by setting a beacon's 'coverage threshold' (P_c) to a higher value.

We next evaluate the variation in coverage at *variable participation rate of users* {100%, 75%, 50%, 25%} throughout each time slot in a day. Figure 6(b) shows comparison of average coverage across all time slots of weekdays for different participation rates. We can see that *Naïve* approach obtains an average of above 90% coverage at all participation rates. This again shows the enough redundancy that exists in user locations at any time of the day. Both the greedy approaches obtain almost similar coverage as that obtained at 100% participation rate. However, in the case of *Instantaneous* approach, rate of drop in coverage is high when the participation rate is reduced. Also, even at lower participation rates, *Instantaneous* approach schedules a large number of users (over 1000), whereas *Greedy* is more robust, scheduling only about 100 users irrespective of participation rates.

2) Fairness: To understand the *fairness* with which phones are scheduled to scan in a day, we first computed the average number of $w = 10$ minutes BLE scans performed by a phone in a day. For comparing the fairness of different approaches in scheduling users to scan, we plotted the cumulative distribution of the numbers of scans in a day for each user.

Figure 7(a) shows the distribution of average number of per-day ($w = 10$ minute long) BLE scans performed by each smartphone. From the figure we can see that the *Greedy* approaches are very fair in terms of scheduling users for scanning as almost all users are selected to scan for only less than 8 slots in a day. However, the *Instantaneous* approach is unfair towards almost 40% of the users who scans for as high as 24 slots (i.e., 4 hours) in a day. We also observed the distribution of the consecutive scan periods, across all users: as expected, the *Energy-Aware Greedy* approach results in a lower value for such consecutive scans (only 12.8% of scanning schedules involved consecutive scans); compared to the *Greedy* approach (where consecutive scans occurred in 17.5% of scanning schedules).

3) Energy-efficiency: Table II shows the comparison of total phone-scans performed over a day by each approach. As expected, the *Naïve* scheduling and *Instantaneous* scheduling

	Naïve	Instantaneous	Greedy	Greedy-Energy
10 minutes phone scans	407376	64659	7724	7658

TABLE II: Comparison of average number of 10 minutes scans performed over a day by all users scanning

approaches incur a very high number of BLE scans in a day. In contrast, the *Greedy* and *Energy-aware greedy* approaches generate only $\sim 10\%$ phone-scans, compared to *Instantaneous* approach.

4) Inter-report Gap per Sensor: We study the frequency of status reports that can be obtained from each of the 2016 sensors present at each landmark. Figure 7(b) plots the CDF of *inter-report gap* per sensor obtained using various approaches. From the figure, we can see that the *Naïve* and *Instantaneous* approaches are able to provide a status report from each sensor at least every 20 minutes, while the *Greedy* approach provides a status report every 20 minutes only for 50% of the sensors (this coverage rate rises to 80% for hourly reports). Certain sensors, at less frequented campus locations, are reported very infrequently using the *Greedy* scheduling approach.

VII. Discussion and Future Work

While our approach can provide frequent updates in densely populated environments, there are several possible ways to extend our current framework.

(1) *Sensors have different criticality* : Currently, we assume that all sensors have the same inter reporting frequency requirement. However, in a practical scenario, the requirement might vary based on sensors and time of day. For example, a 30 minute reporting might be okay for monitoring a dustbin which is placed at a location which is not often visited, but might be too low for a sensor reporting seat availability in a food court during lunch time. In future, we plan to explicitly model such differences in the acceptable *inter-report* gap across sensors.

(2) *Improving Coverage*: Currently, we assume that there is a sensor at every landmark (approx. every 3 meters). Such dense sensor deployment is probably unlikely. Moreover, in cases where we do not get sufficiently frequent reports from a critical section, we can dynamically increase the coverage threshold P_c for the associated sensors, thereby increasing the participation rate in that region.

(3) *Preserving privacy of the mules*: A major concern for participants in location-driven participatory sensing tasks is their individual privacy. *SmartABLE* can be augmented to include techniques such as differential privacy and randomized response, which allow the scheduler to obtain *aggregated* occupancy statistics, without revealing individual movement patterns.

VIII. Conclusion

In this paper, we propose a framework for smart campus monitoring, which leverages on participatory relaying by smartphones to obtain frequent updates from BLE-equipped sensors attached to various indoor resources/facilities. By conducting extensive analysis on location traces from our university campus, we observe that there is sufficient redundancy in user occupancy and propose techniques for collecting such sensor data in an opportunistic, but coordinated, manner. We develop some smart scheduling strategies and compare their relative tradeoffs between coverage and relaying energy cost.

IX. Acknowledgement

This work was supported by Singapore Ministry of Education Academic Research Fund Tier 2 under research grant MOE2011-T2-1001 and by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative. All findings and recommendations are those of the authors and do not necessarily reflect the views of the granting agency, or Singapore Management University.

References

[1] "Estimote beacons," http://estimote.com/, 2012, accessed:2017-09-15.

[2] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Using predictable observer mobility for power efficient design of sensor networks," in *Proc. of IPSN'03*, 2003.

[3] S. Jain, R. C. Shah, W. Brunette, G. Borriello, and S. Roy, "Exploiting mobility for energy efficient data collection in wireless sensor networks," *Mobile Networks and Applications*, 2006.

[4] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2, pp. 215–233, 2003.

[5] U. Park and J. Heidemann, "Data muling with mobile phones for sensornets," in *Proc. of SenSys'11*, 2011, pp. 162–175.

[6] K. Jayarajah, R. K. Balan, M. Radhakrishnan, A. Misra, and Y. Lee, "Livelabs: Building in-situ mobile sensing & behavioural experimentation testbeds," in *Proc. of MobiSys'16*. ACM, 2016.

[7] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: Sensor networks in agricultural production," *IEEE Pervasive computing*, 2004.

[8] O. Tekdas, V. Isler, J. H. Lim, and A. Terzis, "Using mobile robots to harvest data from sensor fields," *IEEE Wireless Communications*, vol. 16, no. 1, pp. 22–28, February 2009.

[9] Y. Qu, K. Xu, J. Liu, and W. Chen, "Toward a practical energy conservation mechanism with assistance of resourceful mules," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 145–158, 2015.

[10] D. Zhao, H. Ma, S. Tang, and X.-Y. Li, "Coupon: A cooperative framework for building sensing maps in mobile opportunistic networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 392–402, 2015.

[11] N. Eagle and A. S. Pentland, "Eigenbehaviors: Identifying structure in routine," *Behavioral Ecology and Sociobiology*, vol. 63, no. 7, pp. 1057–1066, 2009.

[12] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM transactions on networking (TON)*, vol. 19, no. 3, pp. 630–643, 2011.

[13] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky, "Human mobility characterization from cellular network data," *Communications of the ACM*, vol. 56, no. 1, pp. 74–82, 2013.

[14] D. Bhattacharjee, A. Rao, C. Shah, M. Shah, and A. Helmy, "Empirical modeling of campus-wide pedestrian mobility observations on the usc campus," in *Proc. of VTC'04*, vol. 4. IEEE, 2004, pp. 2887–2891.

[15] T. Kandappu, A. Misra, S.-F. Cheng, N. Jaiman, R. Tandriansyah, C. Chen, H. C. Lau, D. Chander, and K. Dasgupta, "Campus-scale mobile crowd-tasking: Deployment & behavioral insights," in *Proc. of CSCW'16*, 2016.

[16] S. Ji, Y. Zheng, and T. Li, "Urban sensing based on human mobility," in *Proc. of UbiComp'16*, 2016.

[17] Y. Liu, B. Guo, Y. Wang, W. Wu, Z. Yu, and D. Zhang, "Taskme: Multi-task allocation in mobile crowd sensing," in *Proc. of UbiComp'16*, 2016.

[18] M. Radhakrishnan, A. Misra, R. Balan, and Y. Lee, "Smartphones and ble services: Empirical insights," in *Proc. of IEEE Mass*. IEEE, 2015.