IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 July 2023, accepted 13 August 2023, date of publication 17 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3306328

RESEARCH ARTICLE

Recognizing Emergencies and Multi-User Behavior Patterns Using Imperfect Data From Distributed Access Points. A Non-Intrusive Proof of Concept

ORESTES G. MANZANILLA-SALAZAR, (Member, IEEE), HAKIM MELLAH, AND BRUNILDE SANSÒ (Senior Member, IEEE)
Electrical Engineering Department, Ecole Polytechnique de Montreal, Montreal, QC H3T 1J4, Canada
Corresponding author: Hakim Mellah (hakim.mellah@polymtl.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant DG 05734.

ABSTRACT This paper presents a privacy-preserving proof of concept for assessing human behavior in emergency scenarios using aggregated data from multiple WiFi access points. The proposed method focuses on preserving individual privacy by avoiding tracking and metadata analysis, while still achieving effective multi-user activity recognition. To implement our approach, raw data from the Eduroam WiFi network at Polytechnique Montreal was collected and analyzed using standard supervised and anomaly detection techniques. The initial test was on recognizing patterns of academic activity, serving as the foundation for our investigation. Subsequently, the same methodology was applied during an evacuation drill scenario to recognize anomaly situations. Through our research, we demonstrate the potential to assess human situations effectively while safeguarding privacy, providing a critical capability for the early detection of emergency situations.

INDEX TERMS Wireless networks, movement patterns, indoor behavior, machine learning, binary classification, anomaly detection, emergency management, data aggregation.

I. INTRODUCTION

Recognizing human behavior is relevant to enable the anticipation and management of human needs. Some behaviors or activities are part of daily or weekly routines, while others are unexpected and possibly of great impact or risk, such as disasters, accidents and violent events. Individual movement can be considered highly independent. However, population movements are highly predictable, and their understanding when in the face of catastrophes is very relevant to manage these situations [1]. The problem of finding patterns in the activities of multiple people, or multi-user activity recognition, however, has been given less attention than individual activity recognition [2]. The study of human movement has been researched via the following:

1) accelerometers or Global Positioning System (GPS) devices,
2) vision-based recognition of human behavior [3], [4],
3) classical radio methods, such as SONAR and RADAR [5],
4) studying the effects of the human body on the physical characteristics of waves (amplitude and frequency) [6].
5) and integrating data from more than one method [7], [8]

Our research addresses methods that do not require specialized sensors that explicitly measure position and movement, such as those in (1). We know that (2) vision-based methods can be affected by light conditions, obstacles, and limits in the field of vision and involve privacy issues. Depending on the Doppler effect, (3) SONAR- and RADAR-like methods are known to be dependent on movement direction and are degraded by the occurrence of multiple paths. Finally,

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

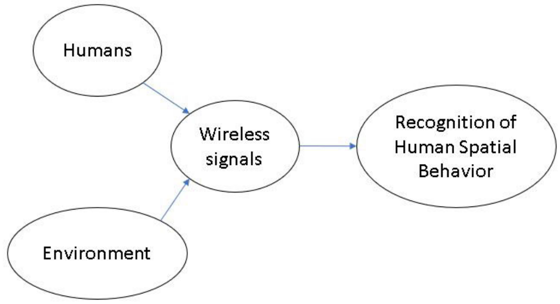

FIGURE 1. Scheme of the approach to use wireless technology to recognize human behavior.

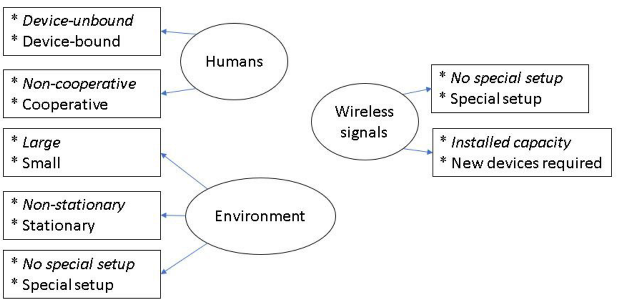

FIGURE 2. Requirements for each subsystem in sensing human behavior via wireless communication signal perturbation.

(4) physical alterations in waves due to human presence is an area of research that has been varied and prolific. An interesting survey of contactless sensing methods used to study human spatial behavior, activities, and even mental health indicators is presented in [9]. The reader is advised to follow the survey presented in [2] to learn about the state of the art in multi-user activity recognition in general.

Due to the focus of this work on wireless technologies, we narrow our analysis to the premises of interest behind such wireless-based techniques, as presented in Figure 1. As depicted, at a higher level, we can describe the process as follows: both humans and the environment in which they dwell affect the way wireless signals are received. As a consequence, perturbations and traces of wireless signals can be analyzed to recognize human behavior. Every wireless-based technique has different requirements from three subsystems: humans, the environment under study, and the wireless network technology generating the signals. In Figure 2, we classify some important aspects of what is required from each of these subsystems.

- *Human subsystem*: Some techniques require some cooperation from humans to train machine learning models and follow specific instructions to enable the methods to work [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Some require humans to carry with them a wireless device or install a specific program on their mobile devices [10], [11], [12], [13], [14], [15], [19], [20], [21], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33].
- *Environment subsystem*: Some techniques require a stationary (or even static) environment [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [28], [29], [30], [31], while some are robust to changes in the environment [27]. Some methods require the environment to be prepared in a particular way to achieve their goals [11], [13], [14], [15], [16], [19], [20], [22], [25], [26], [27], [34], while others can be applied without any special preparation of the environment [18], [21], [23], [24], [28], [29], [30], [31]. Additionally, most human behavior recognition efforts need to be applied indoors and in small controlled spaces.
- *Wireless network*: Some methods require a special setup of wireless devices [10], [11], [12], [13], [16], [18], [22], [23], [24], [25], [26], [27], [29], [30], [31], [34], [35], while others can be used with the wireless infrastructure as it is already installed [19], [28].

In Figure 2, we have placed as the first option for each subsystem, marked in cursive, the option that is less stringent and therefore more desirable. That is, we would prefer methods that can be put in place while people maintain "natural" behavior, using spaces and wireless infrastructure as they are with no extra costs in equipment, and should be applicable in large spaces. Such a combination of requirements, however, is rarely satisfied.

These observations motivated us to pose the following question: *What can be achieved with everyday, basic, widespread WiFi technology (without requiring extra devices, special adaptation of spaces, or active cooperation from humans)?*. Here, we propose an alternative to use these data, with a special interest in unexpected and risky situations, such as disasters or catastrophes. What we explore in this paper is generally aimed at *detecting or identifying events that may simultaneously affect large groups of people in buildings with a single WiFi network*, adding, whenever possible, some extra constraints:

1) Avoid keeping tracking information (to avoid privacy issues).
2) Avoid complex noise filtering (embracing the natural noisy nature of the data and traces).
3) Avoid data cleaning (data coming from devices that are stationary).
4) Avoid complex feature engineering.
5) Avoid sophisticated dimensionality reduction techniques.

In our experiments, we pose the following questions and use machine learning techniques on our data to answer them:

1) Can we discern "routine" behavior patterns of groups of people, such as distinguishing working time from nonworking time? (called Problem I in this paper)
2) Can we use machine learning to detect emergency situations? (called Problem II in this paper)

It is clear that distinguishing working time from non-working time is not a problem that would require any

ML solution. However, we used Problem I to first validate our hypothesis that routine human behaviour can be extracted from very rough wireless information. Once we got good results on Problem I, we decided to set the collection of data for the next evacuation drill.

We are very well aware that evacuation drills do not correspond to real emergencies. Evacuation drills are slow paced events in which people carefully listen to instructions and calmly find the closest exit. However, given the rarity of real emergencies and therefore the difficulty of obtaining real-life data for those events, we believe that the evacuation drill provides an excellent opportunity to study an event that is the closest to an emergency situation. The idea is to create a proof of concept to incite stakeholders to maintain data that maps non-intrusive wireless information to human behaviour so that more refined modelling can be achieved for different types of emergency situations (fire, criminal situation, health emergency, etc.) to have tools for early detection.

Early detection is important to answer as soon as possible the emergency situation and to prevent the creation of stampedes, that although rare, can hinder the well-being of the people involved when they occur.

Our proposed methodology involves aggregating data in time intervals. Other articles have focused on WiFi traces in different ways to find patterns related to human behavior, network usage, or the nature of the usage of the different spaces in the environment. In [36], the traces of connection from mobile phone users to WiFi access points (APs) are analyzed to predict the time that will elapse until they connect to another AP. In [37] and [38], the counts of users connected to each AP are analyzed over time to classify the APs to understand the dynamics of usage that can be expected in each part of the WiFi network. Reference [39] uses user counts per AP to create an interactive 3D visualization of a university campus that shows the size of the crowd present in each building. In [40], users' connections to APs were used to classify APs according to their behavior and to identify periodicity in people's mobility, with the goal of using this information to help create models of human mobility. In [41], the connections of users to each AP are used to model their mobility. The evolution of the number of users in each AP is analyzed over time. The goal is to understand mobility based on real data. In [42], connection logs are analyzed to infer the type of activity that takes place around the APs, which is then validated by comparing the results with information on the usage of each space provided by the university.

None of the aforementioned works, however, addresses the same problem on which we focus in this paper, which is to use very coarse and aggregated WiFi traces to detect human patterns, particularly in the case of emergencies.

To summarize, the main contribution of this article is to propose a methodological proof of concept to detect human activities, including emergency situations, based on coarse wireless data. In particular:

1) We propose a time-aggregation strategy for wireless communication traces that is easy to implement and helps preserve the original shape of the datapoints. We show, for the experiments performed in this work, the impact of each type of aggregating element on the performance of our machine learning techniques. This is a technique that has the potential to be used in other types of problems involving data from wireless networks.
2) We propose a new way to analyze WiFi traces without noise filtering, data cleaning or tracking of individual devices to recognize simultaneous behavior in buildings that provide a widespread WiFi network. This makes the method cheap, easy to implement, and therefore widely applicable in a very short time. We should emphasize, however, that this is limited to behaviors that can be identified due to schedules or planned drills or based on data gathered in previous real occurrences of disasters or catastrophes.

We will now present the problem and our assumptions in the following section.

II. THE PROBLEM

The main assumption of this paper is that we can detect the occurrence of events in spaces served by a wireless network if these events cause large enough groups of people to change the pattern of their ambulatory movements. We focus in particular on investigating the feasibility of doing so based on the footprints that their ambulatory movements make on aggregations on data from WiFi AP logs. We now discuss some of the issues related to the self-imposed stringent constraints on solving the problem.

A. USER PRIVACY

One of the motivations behind the choice of working on aggregated nontracking data is a desire to look for a better solution in the balance between the need for protection of individual privacy and the need for innovation [43]. It is usually accepted that this is a challenging trade-off. Most work done in the recognition of human ambulatory behavior relies on making use of information that is sensitive. Tracking and packet-generation data are particularly suitable to reconstruct even more knowledge about users [42], [44], [45], such as their type of activity, role, type of device, or even identity, which may remain anonymous. For this reason, we propose a method that not only does not track individuals as they traverse the space under study but actually uses aggregations of counts of people, making it unfeasible to reconstruct individual information from the data that is used as input for our aggregation-based preprocessing procedures and our machine learning models and therefore does not violate user privacy.

B. DATA QUALITY CONSTRAINTS

Among the efforts to keep the method as easy to implement as possible, we started our experiments by purposefully skipping several steps that could augment the richness of the data that is used in our models. We also accept the fact that there is valuable information that is technically unfeasible to obtain due to the nature of the method used to gather the data.

For instance, some aspects make our data systematically incomplete, which is expected to make our task harder. In particular, we rely on Access Point (AP) logs, which implies that the behavior of people who are not carrying WiFi-capable and WiFi-active devices is ''invisible'' for our method.

Furthermore, several factors introduce ''noise'' and may hide the behaviors we might be looking for in the data:

- Due to the anonymity, devices returning after disconnecting or being away from the range of the APs are indistinguishable from newly arriving devices.
- Handovers among APs can be triggered by connection-quality problems. We are aware that trying to interpret changes from one AP to another as a consequence of movement therefore might introduce noise to the data.
- Similarly, it is well known that devices can quickly switch between two or more APs, a phenomenon known as the ''ping-pong effect''. Similar to the previous issue, this introduces the illusion of quick movement of a device when it may actually be perfectly static.
- The assumption that human ambulatory behavior is associated with handover among APs is derived from the assumption that the AP that serves a device can somehow be a symbolic representation of its location in space. We acknowledge the fact that the assumption is imperfect, as the AP providing service may not be the closest one to a device.
- We label data based on some predetermined schedules that we expect to have some effect on human behavior (or at least a significant fraction of the people present in the building). This assumption might differ from reality. For example, not every employee follows a predetermined working hours schedule. They may arrive or leave earlier or later than the expected times.

In this paper, we propose an approach to ''live along'' with these imperfect and noisy conditions and investigate whether appropriate preprocessing and use of ML models to recognize the patterns of interest is good enough to obtain useful results.

III. MODELS PROPOSED

In this work, we consider three elements that are modeled: the targeted event, which is assumed to provoke a change in human ambulatory behavior, the network on which we assume said behavior will leave a footprint, and the machine learning models we will use to try predict one based on the other.

A. TARGETED HUMAN BEHAVIOR

For simplicity, we will define ''targeted human behavior'' as the expected change in shared mobility behavior along the space under study by some subset of the population of dwellers as a consequence of an ''event'' of interest. However,

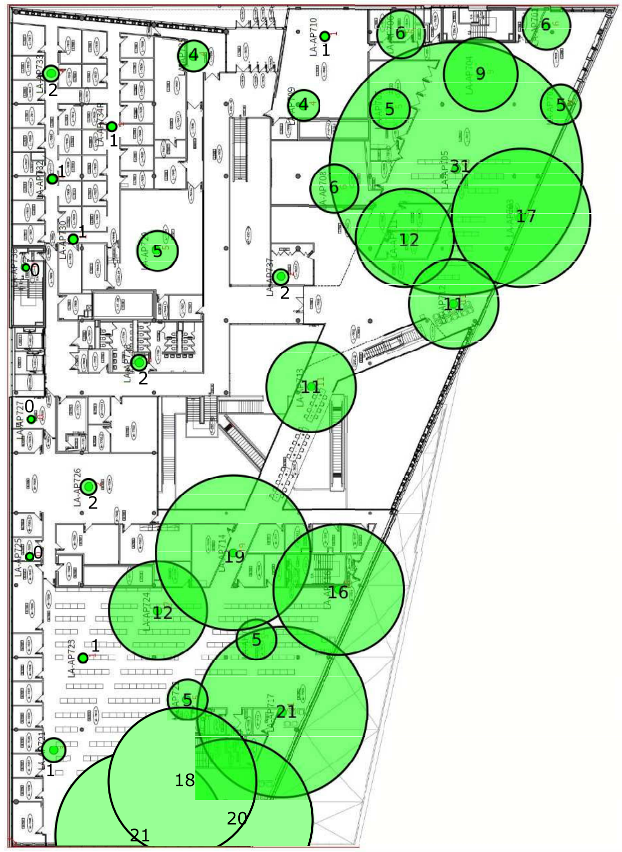

FIGURE 3. Count of devices connected to each AP on one floor of the building.

it is important to keep in mind that said behavior is a proxy to detect an event of interest. We can describe the relevant differences in the types of targeted behaviors as follows:

- *Frequency*: Some may take place as part of an established routine, previously scheduled, while others are rare, unexpected and possibly undesirable, such as those that provoke crises, danger and fear among the dwellers of an indoor space.
- *Spread among dwellers*: The human response to some events may be exhibited by everyone in a building or by some segments of it. Examples of partial spread of the behavior are cases where it is expressed only by those with a specific role, or those in the vicinity of a place, if the eventś effect is local.

In this work, we consider two main problems:

- *Problem I - Detecting working hours*: For this problem, we seek to discriminate if a set of AP logs from a building was gathered during pre-established *working hours*, or not. The targeted human behavior is therefore a frequent behavior but spread only among a subset of the dwellers of a building (i.e., those employees following a work schedule of office hours).
- *Problem II - Detecting evacuation drill*: The event targeted for this problem is the exiting of the building

during an evacuation drill. This behavior is expected to be fully spread among all building dwellers, and it is a rare event. This pattern is of particular interest, as it is a moment in which people behave in a way that is unusual and may be considered similar to that of an emergency situation.

B. THE WiFi NETWORK

Our experiments make use of connection logs from the APs of the *eduroam* wireless network available in one building of Polytechnique Montreal in Canada (Fig. 3). Most devices were connected to the network, but a proportion of them were just in the range of an APs but were not connected, either because their owners did not have the credentials or the setup of the network had not been performed. Each device appearing in the logs received a hash as an identifier to preserve privacy as part of the script run by the daemon.

When pulling the data to apply the methods proposed here, we ignored some information that was available in the sampling with Cisco Connected Mobile Experiences (CMX) application. Specifically, we avoided making use of:

- location estimates for devices,
- APs spatial coordinates,
- signal strength reported for each device, and
- connection status,

C. AGGREGATION MODEL

Data aggregation over time is frequently performed by computing minima, maxima, averages and some measure of dispersion. It is well known that central and dispersion statistics (averages and variances, as typical examples) can have similar values for different original distributions. The aggregation process acts like a filter that hides the shape of the original distribution of the data.

In this paper, we evaluate the effect of different aggregation strategies.

In particular, we tested:

- Baseline aggregation statistics: mean and variance.
- Baseline range measurements: max and mean.
- Percentiles.
- Higher-order momenta: kurtosis and skewness.
- Average estimations of derivatives: first and second.

What we call "first derivative" average estimations consists of computing the average changes in the device counts in consecutive timestamps. A positive value indicates that on average, there is an increase in the device counts, while a negative value indicates the opposite. Similarly, the second derivatives are the average change in consecutive first derivative estimations.

We started using a "baseline" combination of the statistics taken from the device counts (mean and variance) and some specific combinations of these baseline statistics with the others mentioned above. In section V, we will give more details about the sets of statistics that were used in the experimental parameter sweep. We call the set of statistics and values computed to describe the dynamic behavior of the original data simply "statistics" for the sake of brevity.

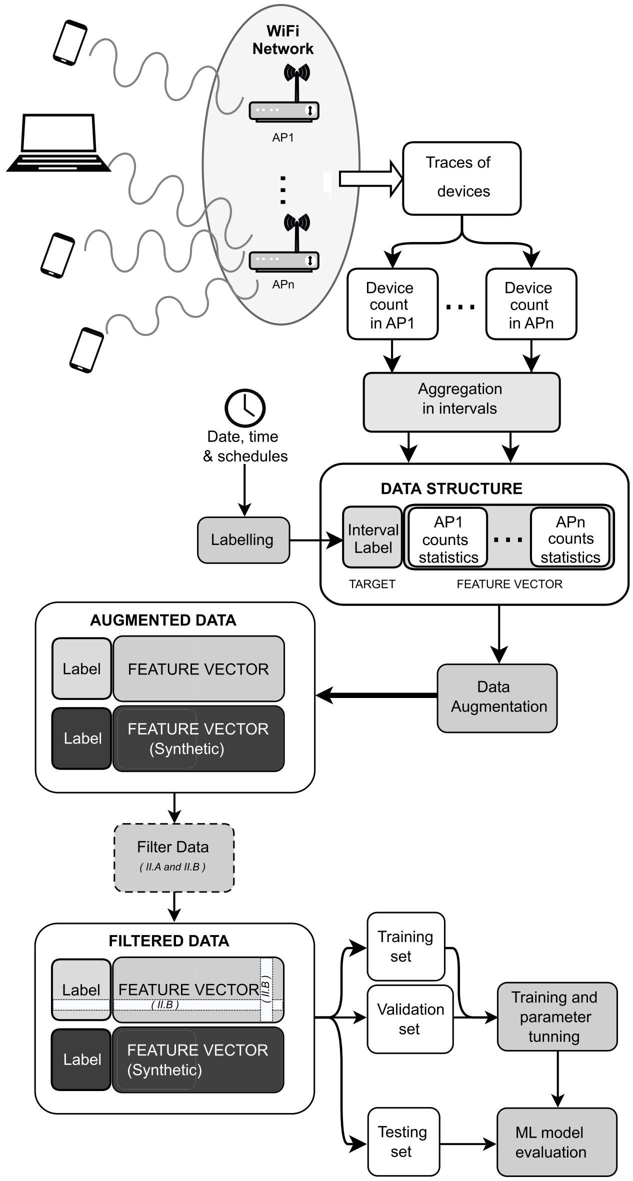

FIGURE 4. Data processing, aggregation and ML training process.

D. MACHINE LEARNING MODELS

We used a binary classification problem approach to tackle problems I and II (see subsection III-A).

Supervised binary classifiers require a pair consisting of (a) input features describing a data point (independent variable) and (b) the target values or label of said data point (dependent variable). After a binary classification model is trained, we expect the output of the model to match (mostly) the target values. In our work, the inputs of a data point (a specific aggregation interval) are the vectors corresponding to the aggregation "statistics" (introduced in subsection III-C and explained in detail in V) of the device counts for each AP.

The labels or target values, on the other hand, depend on the problem we aim to solve, as follows:

- In problem I, which addresses the scheduled working hours, the labels are "nonworking hours" and "working hours".
- In problem II, which addresses detecting the occurrence of the evacuation drill, the labels are "nondrill" and "drill".

In Fig. 4, we represent a simplified scheme of how traces of n APs are processed to obtain the device counts for each timestamp and then how they are aggregated for each interval considered. The resulting data structure for each data point (corresponding to a single time interval) is a vector that adds to the "statistics" of each AP (feature vector) an indicator of the label or target value.

Once we have the features matrix and the labels (targets) vector that will be used to train our Machine Learning (ML) models, we need to address an issue that is present in both problem I and problem II: class imbalance. There are many more data points corresponding to the behavior we want to detect (working hours or the evacuation drill) than those corresponding to moments where those behaviors are not exhibited. This makes any trained binary classifier be biased to give as an output the class corresponding to the majority. This is because guessing the majority class will produce the right answer more frequently. For this reason, the next step in the process described in Fig. 4 is data augmentation.

In our case, both intervals corresponding to "working hours" and "evacuation drills" were a set that contained the minority of the data points. In section IV, we explain in detail how the data were not only gathered, processed, aggregated, labeled and filtered but also augmented.

We trained several binary classification techniques using 10-fold cross-validation to select the best model from a pre-established grid search across the hyperparameters of each model. The best model for each classification technique, fully trained, was then evaluated using data previously separated as the testing set to obtain the performance indicators reported in this paper.

We used the following binary classification model models:

- extra trees,
- logistic regression,
- random forests,
- bagging trees,
- shallow neural networks,
- linear Support Vector Machines (SVMs),
- decision trees,
- gradient boosting trees, and
- naive Bayes.

The performance was evaluated by computing False Positive Rate (FPR), Detection Rate (DR) and Receiver Operating Characteristic (ROC) Area Under the Curve (AUC).

IV. THE DATASET

We now describe the process, which involves:

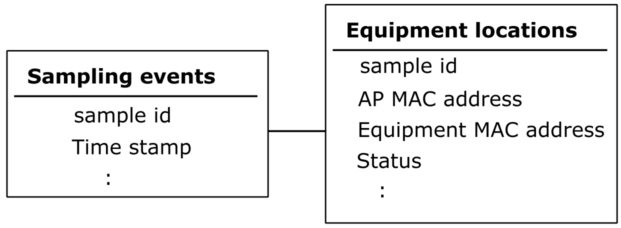

FIGURE 5. Database diagram, with the fields used in our work.

TABLE 1. Typical raw data sampled from AP logs.

Sample ID	UserName	AP MAC Address	Sample time-stamp
1	5bac0b...	e0ed722608...	2017-09-08 15:51:07
1	d704b5...	e934a10b73...	2017-09-08 15:51:07
2	f059ea...	5b50722d57	2017-09-08 15:51:39

1) Data gathering
2) Data aggregation
3) Inclusion of contextual data; and
4) Data augmentation and filtering

A. DATA GATHERING

The WiFi APs logs were gathered using the Cisco CMX monitoring tool, which was used regularly by the network administrator. A daemon was implemented to refresh the web interface to sample the information available from the devices that are detected by each AP. The sampling frequency was approximately two samples per minute. The data included every AP in the building known as Lassonde (see a plan of one floor of the building in Fig. 3) and encompassed a period of 17 days. During this interval, except for a planned evacuation drill, no other unusual event took place at the institution.

Usernames and MAC addresses were assigned a unique hash as soon as they were first detected by an AP of the network (see an example of the raw data in Table 1). The way the script was designed, if a device was detected by an AP again (after leaving the building or being turned off for some reason), it received a new hash. The devices that were gathered in this way included even some that may not have been connected to the WiFi network. This is because the query on the AP logs included those devices that sent a probe frame, a type of packet that is sent when the device attempts to find a nearby network to connect, which can happen even when the WiFi is turned off on laptops and smartphones.

The raw data lacked any information about the type of device (mobile or not) or the type of community member that owned the device (faculty, alumni, employees and visitors), and in our work, no efforts were made to infer or reconstruct this information. Similarly, we did not implement any script to determine whether one device appeared simultaneously in the logs of more than one APs, nor did we estimate in any way the distance from a device to the APs to determine which was the closest.

The network administrator provided the information as a MySQL dump with two tables: one with AP-sampling timestamps and one with the actual traces, linked among themselves via sample IDs (see Fig. 5). In a real-time implementation of the techniques we have developed in this paper, data would need to be fed into a database as a stream. However, such a development was beyond the scope of our work.

B. DATA AGGREGATION

The data used to feed our ML models were not those of the counts. Our approach consisted not of recognizing patterns in a set of data from a single timestamp but instead of aggregations of several timestamps. There are two main reasons to use aggregations:

1) This kind of data is generated as a stream. Therefore, when storing data in its original form or in stamp-by-stamp counts, the data storage requirements can easily become a burden. This is the most frequent motivation behind the aggregations of communications traces.
2) Having statistics and estimates aggregating several traces in a period of time allows us to "see" changes in the distribution between one interval and the other that can be more informative than observing a single instantaneous count. For example, a very high kurtosis indicates that the number of devices is not changing, or a negative estimate of the first derivative indicates that the number of people connected to the AP is increasing.

We experimented with several sizes for aggregation intervals (5, 10 and 15 minutes). For each experimental aggregation size, we ran a script that computed, from the JSON file that contained the device counts, the "statistics" mentioned in subsection III-C describing the distribution and dynamic behavior of the device counts within each interval. The details of the device count "statistics" used to train ML models in each of our experiments are explained in section V. The "statistics" were stored as CSV files, where each row corresponded to the timestamp identifying each aggregation interval, and columns contain the values computed for each AP.

C. INCLUSION OF CONTEXTUAL DATA

We added labels to our CSV files with aggregated data columns that indicated the patterns we wanted to be able to recognize. We also included additional contextual data in extra columns, which helped us filter the data (data filters used are explained in subsection IV-D).

We added to each row of the CSV files:

- A label that indicated whether the evacuation drill was taking place (specifically, the moments in which people were exiting the building, not including their return).
- Whether an interval corresponded to the start (first 30 min) of the lunch hour scheduled in the working hours for administrative staff.
- Whether an interval corresponded to the end of the working hours (first 30 min after the working hours ended).

To decide whether an interval of aggregation can be labeled as belonging to the set corresponding to one of the target behaviors (working hours or evacuation drill), our script compares the start and end timestamps of the interval being aggregated, with the start and end of the interval in which the target ambulatory behavior is expected to happen. If the two overlap, the label corresponding to the behavior is applied.

D. DATA AUGMENTATION AND FILTERING

As mentioned in subsection III-D, the two problems addressed presented class imbalance. There are two main sampling approaches to tackle the problem of class imbalance: undersampling, which in our case implies losing most of our data, and oversampling. We chose to implement a strategy in which we synthetically oversampled the minority using the Synthetic Minority Oversampling Technique (SMOTE) because it creates plausible synthetic minority class samples. Samples generated via SMOTE are considered "plausible" in the sense that they are similar, or close in feature space, to the rest of the minority set data points and therefore are widely used and assumed to be better than simple random oversampling.

After the data were augmented, two types of "filtering" were used: (a) we used Principal Component Analysis (PCA) on the data, which can be considered a statistical filter that reduces the dimensionality of the data, and (b) actual filtering by columns or rows. The motivation for actually filtering the dataset is an experimental decision that is mentioned in section VI after performing our base experiments with problems I and II.

As shown in Fig. 4, after the data were augmented and filtered, they were separated into training (10-fold cross-validation) and testing sets.

V. EXPERIMENTAL SETUP

Our main goal is to test our approach to find shared human ambulatory patterns of interest. One type of pattern of interest is "massive fleeing", which may be associated with risky or dangerous events such as explosions, catastrophes or terrorist attacks. The WiFi logs of APs of a building during periods that contained both "mundane" routine behaviors as well as a drill exercise share some characteristics with the type of risky patterns in which we are interested. Our experimental setup consisted of starting with a "mundane" pattern, in this case, our problem I: that of detecting how people who follow an office-hours schedule are actually in working, solely based on the fingerprints obtained from the WiFi logs. As will be detailed in section VI, this task was performed successfully.

We then proceeded to use our method for our evacuation drill detection task, problem II.

We tried different aggregation sizes and different combinations of "statistics", experimented with and without PCA, and tried several ML models to determine what conditions would give the best performance in discriminating the patterns of interest of each problem and subproblem.

We performed 2377 experiments, produced by sweeping across the following "levels" for each of the experimental parameters:

- 9 binary classification models;
- Use or not of PCA;
- 3 different sizes of the aggregation intervals: 5, 10 and 15 min;
- 11 different combinations of statistics of the device counts; and
- 4 tasks (problems I and II and two additional subproblems that were motivated by some of our results).

The 11 combinations of statistics used for data aggregation are as follows:

1) Mean and variance,
2) mean, variance, minimum and maximum,
3) mean, variance and five percentiles (0.05, 0.25, 0.50, 0.75, 0.95),
4) mean, variance and kurtosis,
5) mean, variance and skewness,
6) mean, variance and an estimate of the first derivative,
7) mean, variance and an estimate of the first and second derivatives,
8) mean, variance, minimum, maximum and five percentile ranges,
9) mean, variance, kurtosis and skewness,
10) mean, variance, min, max, percentiles, kurtosis and skewness.
11) mean, variance, min, max, percentiles, kurtosis, skewness, and first and second derivatives.

In this paper, we refer to a specific combination of levels in each experimental parameter as an "experimental condition".

VI. ANALYSIS OF RESULTS

Most of our results were good. The fact that a large majority of the experimental conditions were very similar made it difficult to discern which was the best "level" for each experimental parameter by simple comparison of ROC AUC, FPR or DR values. To determine which experimental condition level was the best, instead of averaging the effect of a single condition among several levels of the other experimental conditions, we took what we call a "popularity" approach. The procedure followed to determine "popularity" is as follows:

1) Sorting all the experimental results according to the classification performance (ROC AUC);
2) Analyzing the sorted list of results and selecting the top-performing experimental conditions; and
3) Computing the frequency of each experimental parameter "level", from this set of top performers (calling the more frequent parameter levels the top performers).

In the scenarios where there was extremely low variability among the values of ROC AUC, the procedure was performed by sorting based on FPR. The procedure was repeated by observing the frequency of each of the experimental parameter levels when taking the low end of the sorted list (the "bottom"-performing experiments). We looked for consistency in experimental parameter levels based both on their *more frequent* occurrence in the top-performing set and their *less frequent* occurrence in the bottom-performing set. When both things happened, we said the experimental parameter level was "popular".

This approach shows clearer patterns and interactions of experimental parameter levels than making statistical performance average comparisons.

A. STRESSING THE METHOD

Because our results for problem II were particularly good (FPR was never higher than 0.5 %, for any of the experimental conditions), we designed two new experiments to stress the method and explore its limitations: II-A and II-B. Both subproblems implied limiting (filtering) the information that was fed to the ML models in two ways (depicted in Fig. 4 after the step of data augmentation):

1) For subproblem II-A, we filtered to randomly eliminate columns, effectively omitting most of the features used in the problem.
2) For subproblem II-B, we filtered rows, leaving only those of the majority class that were expected to be most similar to the minority class.

The first derived subproblem, which we call problem II-A, repeated the experiments while taking the device counts of only a randomly chosen 10 % of the APs. The second derived subproblem, called problem II-B, consisted of limiting data from "normal" intervals (not during the evacuation drill) so that we considered only those where a large proportion of people with a pre-fixed schedule were expected to leave the building at the same time (namely, at the start of the lunch hour and the end of the working hours). Problem II-B, therefore, only kept "normal" behaviors that were expected to be the "most similar" to what would happen during the evacuation drill exercise, under the assumption that the task of discriminating between the two patterns would be more difficult.

B. THE SEARCH FOR THE "BEST" EXPERIMENTAL CONDITIONS

In Table 2, we present the patterns observed for the "popularity" of each of the experimental conditions. With respect to the statistics used (first row of Table 2), for problems I and II-A, we could not find a clear preference. All the combinations of statistics used seem to produce excellent results. For problems II and II-B, however, we observed some preferences. In problem II, the most present among the best results were those that included only 1 or 2 shape-related statistics (such as asymmetry and kurtosis), while in problem II-B, those sets that involved the most complex statistics along with minimum and maximum were the most frequently found among the best results.

There was a clearer pattern when observing the "popularity" of ML models (second row of Table 2). While, for problem II-B, all binary classifiers were present among the best results in similar quantities, for problems I, II and II-A, extra trees were systematically the most popular (especially for problem II). In the table, we present the three most frequent, in order of their frequency (in descending order).

The size of the aggregation intervals (third row of Table 2) was mostly nonconclusive, except for problem II-A, where the 15-minute aggregations were more frequent among the best-performing experiments. Similarly, the use of PCA (fourth row of Table 2) presented no significant preference at all.

In Table 3, we show our findings for those experimental conditions that presented interesting interactions. When observing the joint effect of the models used versus the use or not of PCA for dimensionality reduction (first row of Table 3), we found that some models (which are not the most "popular" ones) were more popular when used with PCA (ANNs, logistic regression, random forests and bagging trees). For subproblems II-A and II-B, only ANNs seemed to be more popular when PCA was used.

Analyzing the "popularity" of the experiments under interactions between the sets of statistics used and the binary classifier used for pattern discrimination (second row of Table 3), we found that extra trees, the predominant model, along with logistic regression, performed perfectly for problem I when used with any of the sets of statistics. For problem II, we observed that all the models benefited from more "sophisticated" statistics (i.e., those beyond simply using mean and variance). Interestingly, for problem II-A, using only 10 % of APs, extra trees had systematically perfect results for every set of statistics used. For problem II-B, the only evident systematic interaction between model choice and statistics used was that ANNs performed better if 2 or more complex statistics were used (beyond mean and variance).

There are three experimental conditions for which there are no clear recommendations: choice of statistics, size of the aggregations, and use of PCA. However, given that extra trees were consistently and frequently included among the sets of best experiments, we can assign it as the model of choice for our data in all our problems and again study the statistics, aggregation size and use of PCA. Additionally, given that problem II and its subproblems II-A and II-B are extremely easy to solve, we can focus on the best levels for aggregation size and PCA for problem I.

In problem I, we averaged across the three aggregation sizes to more closely observe the performance obtained for each set of statistics with or without PCA. We can see that the set of statistics # 3 (baseline and kurtosis) is the best one when using PCA and the second-best one without using it (see Fig. 6). However, without PCA, the best set of statistics (which includes the baseline, minimum and maximum, percentiles, kurtosis and skewness) also has a high value of variance in the ROC AUC (among the 3 different aggregation sizes), which is not the case with the set of statistics # 3. Therefore, we consider that the more consistently well-performing set of stats, with or without PCA, is set # 3.

Regarding PCA, in Fig. 6, it is clear that with extra trees, not using PCA is systematically better. Before jumping to conclusions regarding PCA, we decided to study PCA better by committing not only to the use of extra trees but also to the set of statistics # 3. This resulted in the performance values depicted in Fig. 7, in which there are no averages taken. In this figure, we can confirm that not using PCA is the best option across all interval sizes tested. Interestingly, the gap between the performance values obtained using PCA seems to diminish as the size of the aggregation intervals increases. The best performance for problem I is definitely obtained using extra trees without PCA. Aggregations of 5 and 15 minutes give almost identical performance, with 15 being the best one.

C. RECOMMENDATIONS TO DISCRIMINATE WORKING HOURS PATTERN

When trying to detect whether an interval corresponds to working hours, the best solution is to:

- Use extra trees.;
- Use the mean, variance and kurtosis of the device counts for each AP;
- Not use PCA; and
- Aggregate data in intervals of 15 minutes.

In our experiences, this achieves:

- ROC AUC: 0.974
- FPR: 0.00 %
- Detection Rate: 94.83 %

We conclude with this base problem that the general task of recognizing shared human behavior is feasible.

D. RECOMMENDATIONS TO DISCRIMINATE THE DRILL PATTERN

When trying to detect whether an interval encompasses the moments in which the evacuation drill was in process, the best solution is to:

- Use extra trees;
- Use mean and variance (all sets of statistics work, so using the simplest one seems best);
- Use PCA optionally (i.e., its use has no impact); and
- Aggregate data in intervals of 5 minutes (the use of other intervals makes no difference, but being able to discern from data of smaller intervals is best).

In our experiences, this achieves:

- ROC AUC: 1.000
- FPR: 0.00 %
- Detection Rate: 100.00 %

We should note that these recommendations hold for the more challenging problems of achieving the task with:

- only 10 % of the APs (problem 2-A), and
- comparing only with "normal leaving" behavior (problem 2-B).

After looking at these results, a natural question emerges: *Why is it easier to detect the pattern produced by the evacuation drill than the "baseline" problem of detecting working*

TABLE 2. Effects of experimental conditions.

Experimental parameter	Problem I	Problem II	Problem II-A	Problem II-B
Statistics set	No clear pattern	Best ones include 1 or 2 shape-related statistics. ''More is not always better''	No clear pattern	The best ones are the two more complex ones, along with Baseline + min/max
ML models	1. Extra trees 2. Logistic Regression 3. Random Forests	1. Extra Trees (FPR never higher than 0.5% 2. Naive Bayes 3. Logistic Regression	1. Extra Trees 2. Random Forests 3. Bagging Trees	All models perform well. The lower performance was given by ANNs
Aggregation interval size	No clear pattern		The longer, the better	No clear pattern
PCA	No clear pattern			

TABLE 3. Interactions among experimental conditions.

Interaction	Problem I	Problem II	Problem II-A	Problem II-B
Model vs PCA	PCA is required to get better results in ANNs and logistic regression	PCA is required to get better results in ANNs, random forests and bagging trees	There are good results both with and without PCA (ANN particularly improves with PCA)	No preference in general regarding PCA. ANNs still rely heavily on PCA
Statistics used vs PCA	Extra trees and logistic regression have excellent results (FPR = 0 and high ROC AUC) with all sets of statistics	Addition of more and more ''sophisticated'' statistics is generally good	Extra trees had perfect results (ROC AUC = 1) with every set of statistics	No preference in general. ANNs do better more frequently with the 2 more complex sets of stats

hours? We provide some insight into this question in the discussion section.

VII. DISCUSSION

A. WHY IS PROBLEM II EASIER?

Initially, in our experimental design, detecting a ''normal'' (routine) human behavior pattern, such as the behavior of people following a fixed working schedule, was considered a ''baseline'' problem. This is because it is a common and frequent behavior, thus providing a larger amount of available data that represents that particular pattern. Recognizing working hours proved to be a more challenging problem, with lower performance in classification, compared to recognizing behavior during an evacuation drill.

By intuition, we have a sense that the following reasons contribute to the observed differences.

1) The behavior exhibited during the drill is more distinct compared to how people respond to their working schedules. When the alarm is triggered, individuals are likely to respond immediately, not lingering in their current activities to complete work tasks.
2) The drill behavior is shared by all occupants of the building, which means there is no ''noise'' associated with individuals not adhering to this behavior. In contrast, when studying Problem I, we analyze device counts where most of the active WiFi-capable mobile devices do not belong to people following a strict working hours schedule (such as students, faculty, and visitors). Therefore, Problem I involves more noise.
3) During the drill, as it is a behavior affecting all occupants, there is a higher density of people moving towards the exits. This higher density increases the likelihood of individual behaviors matching those of others. Moreover, individuals tend to exhibit more similar movement speeds during the drill. In ''normal'' situations, people moving around the building may

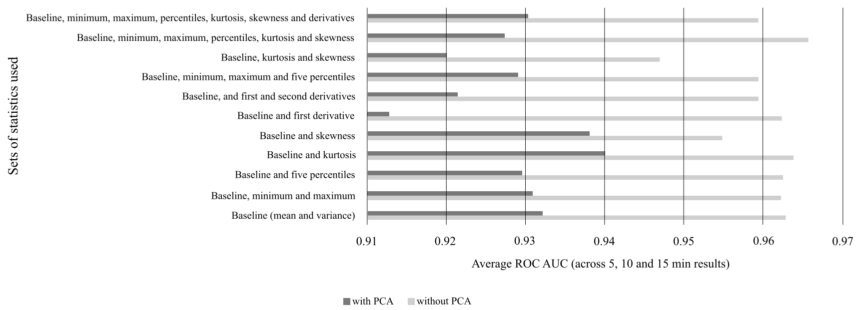

FIGURE 6. Interaction between statistics used and PCA with extra trees in problem I.

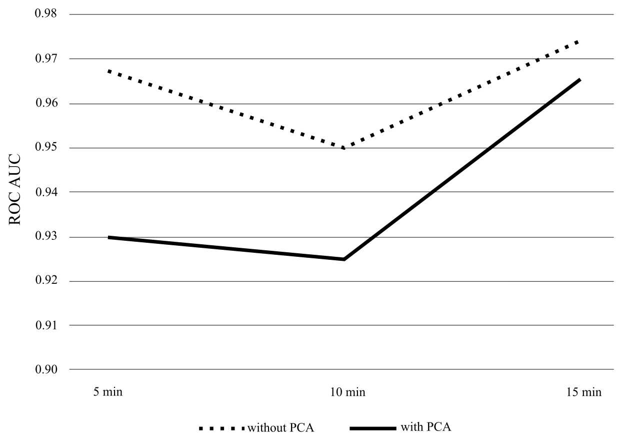

FIGURE 7. Interactions between size of aggregation intervals and PCA, when using mean, variance and kurtosis as aggregation statistics, with extra trees in problem I.

walk at varying distances from one another and display more diverse speed patterns.

4) Furthermore, since the drill schedule is unknown, individuals cannot anticipate it and leave earlier as they would for lunch or at the end of the day.

By considering these factors, one can see why Problem II is comparatively easier than Problem I. Although it is not straightforward to sort those factors in order of importance, our understanding is that the "noise" factor is determinant.

B. EXTENDABILITY OF THE METHOD

The experiments for this study were conducted in an indoor campus setting. We would like to address the extendability of this method to other indoor or outdoor environments.

In terms of indoor settings, such as commercial malls, hospitals, government buildings, airports, and railroad/metro stations where Wi-Fi is available, there are no inherent limitations preventing the implementation of this method. Additionally, the method can be effectively applied in single-floor settings, offering even better detection of emergencies and unusual events compared to our study where AP device counts could span several floors.

Another question that arises is whether a similar method can be implemented using a different wireless technology. Our intuition suggests that if the alternative technology offers base station coverage with a granularity similar to indoor Wi-Fi, then the method would be viable. However, if the technology has a larger base station coverage, device counting alone may not provide sufficient information.

This brings us to the issue of outdoor deployment. Our methodology relies on device counts to extract statistics that can be correlated with user displacement, indicating emergency situations. It is important to note that indoor confinement naturally restricts and directs human mobility, which aids in detecting anomaly situations. Thus, in an outdoor setting, we would lack this advantageous constraint provided by indoor premises.

Furthermore, when considering outdoor coverage, current cellular technology has a much larger coverage granularity compared to indoor Wi-Fi. Consequently, it becomes unclear what types of emergencies could be detected solely through device counting. However, our intuition suggests that macro events spanning multiple areas of a city could still be observed using this approach, although detecting problems within a small area would be challenging.

Nonetheless, as outdoor cells become smaller and their coverage is reduced, more information can be derived from device counting. In such cases, the methodology presented here could potentially assist in early emergency detection.

VIII. CONCLUSION

This paper demonstrates the effectiveness of our proposed techniques, which rely solely on device counts at each access point (AP), in detecting events that result in significant changes in ambulatory patterns of large groups of people in wireless network spaces. Our method requires no tracking, collaboration, or meta-data analysis, imposes no constraints, and utilizes existing hardware and software. Among the classification techniques tested, extra trees emerged as the most suitable. We achieved excellent performance in detecting emergency-like patterns with low false positives. The technique can be adapted for different behavioral groups and localized emergencies. We urge stakeholders to continuously collect and analyze device counting logs for early detection and management of emergencies. While the methodology is technology-independent and can be adapted by mobile service providers, further research is needed to determine its effectiveness in different outdoor settings.

ACKNOWLEDGMENT

This work was made possible by the computational resources of Compute Canada. The authors thank Ludovic Bilodeau-Laflamme and Jesus A. Sanchez-Davila for their assistance with computer implementations. They are grateful to the Computer Service of Polytechnique Montréal for allowing the captures of data from its EDUROAM network. They also acknowledge in particular the work of Lionel Valero who set up the capturing and anonymizing filters.

REFERENCES

[1] V. D. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," *EPJ Data Sci.*, vol. 4, no. 1, p. 10, Dec. 2015.

[2] Q. Li, R. Gravina, Y. Li, S. H. Alsamhi, F. Sun, and G. Fortino, "Multi-user activity recognition: Challenges and opportunities," *Inf. Fusion*, vol. 63, pp. 121–135, Nov. 2020.

[3] L. Wang, G. Zhao, L. Cheng, and M. Pietikäinen, *Machine Learning for Vision-Based Motion Analysis: Theory and Techniques*. Cham, Switzerland: Springer, 2010.

[4] L.-F. Wu, Q. Wang, M. Jian, Y. Qiao, and B.-X. Zhao, "A comprehensive review of group activity recognition in videos," *Int. J. Autom. Comput.*, vol. 18, no. 3, pp. 334–350, Jun. 2021.

[5] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar Navigat.*, vol. 11, no. 1, pp. 107–115, Jan. 2017.

[6] A. Khalili, A. Soliman, M. Asaduzzaman, and A. Griffiths, "Wi-Fi sensing: Applications and challenges," *J. Eng.*, vol. 2020, no. 3, pp. 87–97, 2020.

[7] A. De San Bernabe, J. R. Martinez-de Dios, and A. Ollero, "Efficient integration of RSSI for tracking using wireless camera networks," *Inf. Fusion*, vol. 36, pp. 296–312, Jul. 2017.

[8] J. Cao, W. Li, C. Ma, and Z. Tao, "Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition," *Inf. Fusion*, vol. 41, pp. 68–79, May 2018.

[9] M. Nouman, S. Y. Khoo, M. A. P. Mahmud, and A. Z. Kouzani, "Recent advances in contactless sensing technologies for mental health monitoring," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 274–297, Jan. 2022.

[10] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[11] N. Patwari and J. Wilson, "Spatial models for human motion-induced signal strength variance on static links," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 791–802, Sep. 2011.

[12] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 1230–1239.

[13] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 1–10, Sep. 2017, doi: 10.1145/3078855.

[14] H. Yigitler, O. Kaltiokallio, R. Hostettler, A. S. Abrar, R. Jäntti, N. Patwari, and S. Särkkä, "RSS models for respiration rate monitoring," *IEEE Trans. Mobile Comput.*, vol. 19, no. 3, pp. 680–696, Mar. 2020.

[15] F. Zhang, D. Zhang, J. Xiong, H. Wang, K. Niu, B. Jin, and Y. Wang, "From Fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–12, Mar. 2018, doi: 10.1145/3191785.

[16] M. Scholz, T. Riedel, M. Hock, and M. Beigl, "Device-free and device-bound activity recognition using radio signal strength," in *Proc. 4th Augmented Hum. Int. Conf.*, New York, NY, USA, Mar. 2013, pp. 100–107, doi: 10.1145/2459236.2459254.

[17] S. Sigg, U. Blanke, and G. Tröster, "The telepathic phone: Frictionless activity recognition from WiFi-RSSI," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2014, pp. 148–155.

[18] G. Cohn, D. Morris, S. N. Patel, and D. S. Tan, "Your noise is my command: Sensing gestures using the body as an antenna," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2011, pp. 791–800, doi: 10.1145/1978942.1979058.

[19] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38, doi: 10.1145/2500423.2500436.

[20] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.

[21] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, New York, NY, USA, 2018, pp. 401–413, doi: 10.1145/3210240.3210335.

[22] Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in *Proc. 7th Int. Conf. Mobile Ad-hoc Sensor Netw.*, Dec. 2011, pp. 138–145.

[23] I. Sobron, J. Del Ser, I. Eizmendi, and M. Vélez, "Device-free people counting in IoT environments: New insights, results, and open challenges," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4396–4408, Dec. 2018.

[24] Q. Xu, Y. Chen, B. Wang, and K. J. R. Liu, "TRIEDS: Wireless events detection through the wall," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 723–735, Jun. 2017.

[25] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.

[26] S. Sigg, S. Shi, and Y. Ji, "RF-based device-free recognition of simultaneously conducted activities," in *Proc. ACM Conf. Pervasive ubiquitous Comput. Adjunct Publication*, New York, NY, USA, Sep. 2013, pp. 531–540, doi: 10.1145/2494091.2497611.

[27] M. Reschke, S. Schwarzl, J. Starosta, and S. Sigg, "Situation awareness based on channel measurements," in *Proc. IEEE 73rd Veh. Technol. Conf.*, May 2011, pp. 1–5.

[28] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, Jul. 2015.

[29] H. Lv, M. Liu, T. Jiao, Y. Zhang, X. Yu, S. Li, X. Jing, and J. Wang, "Multi-target human sensing via UWB bio-radar based on multiple antennas," in *Proc. IEEE Int. Conf. IEEE Region 10*, Oct. 2013, pp. 1–4.

[30] M. Arai, H. Kawamura, and K. Suzuki, "Estimation of ZigBee's RSSI fluctuated by crowd behavior in indoor space," in *Proc. SICE Annu. Conf.*, Aug. 2010, pp. 696–701.

[31] F. Zhang, C. Chen, B. Wang, and K. J. Ray Liu, "WiSpeed: A statistical electromagnetic approach for device-free indoor speed estimation," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2163–2177, Jun. 2018.

[32] A. Bugarın and J. M. Alonso, "Estimation of customer activity patterns in open malls by means of combining localization and process mining techniques," in *Proc. Workshop Phys. Agents*, vol. 1285. Cham, Switzerland: Springer, 2021, p. 30.

[33] Q. Chen, I.-T. Chuang, and A. Poorthuis, "Entangled footprints: Understanding urban neighbourhoods by measuring distance, diversity, and direction of flows in Singapore," *Comput., Environ. Urban Syst.*, vol. 90, Nov. 2021, Art. no. 101708. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0198971521001150

[34] N. Thakur and C. Y. Han, "An ambient intelligence-based human behavior monitoring framework for ubiquitous environments," *Information*, vol. 12, no. 2, p. 81, Feb. 2021. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/81

[35] L. Li, X. Chen, Q. Li, X. Tan, J. Chen, D. Wang, and P. Jia, "Contextualizing human dynamics: Understanding the semantics of movement trajectories with Wi-Fi data," *Travel Behav. Soc.*, vol. 25, pp. 183–192, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214367X21000673

[36] J. Manweiler, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Predicting length of stay at WiFi hotspots," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 3102–3110.

[37] F. Calabrese, J. Reades, and C. Ratti, "Eigenplaces: Segmenting space through digital signatures," *IEEE Pervasive Comput.*, vol. 9, no. 1, pp. 78–84, Jan. 2010.

[38] S. Chaikul, S. Phithakkitnukoon, and C. Ratti, "Xplaces: Segmenting physical space through Wi-Fi traces using eigendecomposition and X-means," *IEEE Access*, vol. 10, pp. 17377–17388, 2022.

[39] A. Binthaisong, J. Srichan, and S. Phithakkitnukoon, "Wi-Crowd: Sensing and visualizing crowd on campus using Wi-Fi access point data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, New York, NY, USA, Sep. 2017, pp. 441–447, doi: 10.1145/3123024.3124413.

[40] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Pers. Ubiquitous Comput.*, vol. 11, no. 6, pp. 465–479, Aug. 2007.

[41] M. Kim and D. Kotz, "Modeling users' mobility among WiFi access points," in *Proc. Workshop Wireless Traffic Meas. Model.*, 2005, pp. 19–24.

[42] G. Poucin, B. Farooq, and Z. Patterson, "Activity patterns mining in Wi-Fi access point logs," *Comput., Environ. Urban Syst.*, vol. 67, pp. 55–67, Jan. 2018.

[43] E. Horvitz and D. Mulligan, "Data, privacy, and the greater good," *Science*, vol. 349, no. 6245, pp. 253–255, Jul. 2015.

[44] A. J. Ruiz-Ruiz, H. Blunck, T. S. Prentow, A. Stisen, and M. B. Kjærgaard, "Analysis methods for extracting knowledge from large-scale WiFi monitoring to inform building facility planning," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2014, pp. 130–138.

[45] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: Using social network as a side-channel," in *Proc. ACM Conf. Comput. Commun. Secur.*, Oct. 2012, Art. no. 628637.

ORESTES G. MANZANILLA-SALAZAR (Member, IEEE) received the B.S. degree in production engineering and the M.S. degree in systems engineering from Simón Bolívar University (USB), Venezuela, in 2002 and 2005, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with Ecole Polytechnique de Montreal, Montreal, QC, Canada. From 2008 to 2015, he was an Assistant Lecturer with USB. He taught courses in the areas of mathematical programming, decision science, and operations management, and focused his research on linear and integer programming models to create supervised classifiers. He worked for six years in several positions for private firms in the areas of engineering, technology, and finance. His research interests include machine learning and analytics tools for telco network management and network-enabled systems.

HAKIM MELLAH received the B.Eng. and M.A.Sc. degrees in electrical engineering and the Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 1996, 1999, and 2018, respectively. He is currently a Postdoctoral Researcher with Ecole Polytechnique de Montreal. His research interests include machine quality of experience for the IoT applications in a smart-city environment.

BRUNILDE SANSÒ (Senior Member, IEEE) is currently a Full Professor of telecommunication networks with the Department of Electrical Engineering, Ecole Polytechnique de Montreal, and the Director of the LORLAB, a research group dedicated to developing methods for the design and performance of wireless and wireline telecommunication networks. Over her career of more than 25 years, she has received several awards and honors, has published extensively in the telecommunication networks and operations research literature, and has been a consultant for telecommunication operators, equipment and software manufacturers, and the mainstream media.