Enabling Identification and Behavioral Sensing in Homes using Radio Reflections

Chen-Yu Hsu, Rumen Hristov, Guang-He Lee, Mingmin Zhao, Dina Katabi
MIT CSAIL
{cyhsu,rhristov,guanghe,mingmin,dk}@mit.edu

ABSTRACT

Understanding users' behavior at home is central to behavioral research. For example, social researchers are interested in studying domestic abuse, and healthcare professionals are interested in caregiver-patient interaction. Today, such studies rely on diaries and questionnaires, which are subjective, erroneous, and hard to sustain in longitudinal studies. We introduce Marko, a system that automatically collects behavior-related data, without asking people to write diaries or wear sensors. Marko transmits a low power wireless signal and analyses its reflections from the environment. It maps those reflections to how users interact with the environment (e.g., access to medication cabinet) and with each other (e.g., watch TV together). It provides novel algorithms for identifying who-does-what, and bootstrapping the system in new homes without asking users for new annotations. We evaluate Marko with a one-month deployment in six homes, and demonstrate its value for studying couple relationships and caregiver-patient interaction.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; • **Computing methodologies** → *Neural networks*;

KEYWORDS

Behavioral Sensing, Passive RF Sensing, User Identification, Wireless Sensing, Smart Environments.

ACM Reference Format:
Chen-Yu Hsu, Rumen Hristov, Guang-He Lee, Mingmin Zhao, Dina Katabi. 2019. Enabling Identification and Behavioral Sensing, in Homes using Radio Reflections. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3290605.3300778

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300778

1 INTRODUCTION

Learning users' behavior in their homes is central to behavioral research. Social researchers are interested in how different family members spend their time at home, and how they interact [13, 21, 26, 40, 44]. For example, knowing how a couple balances their work hours to spend time together helps us understand their relationship. Healthcare professionals would like to understand patients' health conditions and caregiver-patient interaction at home [14, 15, 19, 29, 55]. For example, do the patients take their medications at the prescribed times every day? Are the patients independent for most activities of daily living or do they need frequent help from their caregivers? Knowing the answers to such questions helps doctors deliver better care, and guide decisions regarding home-stay versus moving the patient to a nursing home [15, 29, 55].

Current solutions to in-home behavioral sensing rely mainly on self-reporting, i.e., having the subjects write diaries or answer questionnaires. These techniques, however, are often prone to subjective biases and inaccuracies [39, 60]. Moreover, keeping a detailed diary or administering questionnaires incurs a significant overhead from the subject and is not sustainable in long-term studies. An alternative approach would use wearable sensors or smartphones for behavioral sensing [6, 12, 32, 46, 56]. However, older adults could feel encumbered by wearables and uncomfortable using them [17, 54]. Further, past studies have shown that wearable devices lead to adherence problems because people stop using the wearable sensor with time [10, 16, 33, 34, 51].

Our goal is to provide a tool that minimizes the overhead associated with in-home longitudinal behavioral studies. We would like to use the radio (RF) signals that bounce off people's bodies to enable behavioral sensing at home, without diaries or wearables. We build on past advances in passive wireless localization, which transmit a low-power wireless signal and use its reflections to localize people in the vicinity of the radio [4, 27, 37]. Our intuition is that in-home location embeds a wealth of information about user behavior. For example, if a nurse reaches out to the medication cabinet then approaches the patient, she is likely giving him his medications. Similarly, if a person reaches out to the fridge, it indicates a desire to eat. By continuously tracking people's trajectories at home and tagging those trajectories with user identities, one can develop a passive tool for in-home social and behavioral studies. Such a tool incurs no overhead from the users, who continue living their lives without having to wear sensors, answer questionnaires, or write diaries. It is also minimally invasive in comparison to other passive solutions, e.g., cameras and microphones. Even in scenarios where the tool does not provide all data necessary for the study, the tool may still be used to augment and correct self-reported data, which is known to be incomplete and error-prone [39, 60].

Developing such a tool is challenging. First, passive RF localization systems cannot tell the identity of the monitored person. As a result, measurements from different people are entangled, making it difficult to provide meaningful insights. While there has been some initial effort to add identity to such systems [3, 9, 23, 50, 57, 62, 64], none of the proposed solutions work in the presence of multiple people (which is typical in homes). Further, they usually restrict users to walk along one or a few predetermined lines without stopping, and fail when users do not follow instructions [9, 57].

To address this challenge we introduce a new algorithm that combines information over space and time, and operates over both location data and raw RF signals. Our algorithm uses continuity in space and time to stitch location measurements into short trajectories, where each trajectory tracks the motion of one person. It uses these trajectories to spatially separate the received RF signal into multiple signal components, where each component captures the RF reflections from one person along one trajectory. The algorithm provides each trajectory and the corresponding RF component to a convolutional neural network (CNN) trained to recognize user identity. The algorithm works in the presence of multiple people, and has no assumption on user motion; hence it works in the wild while users go about their lives.

Another important challenge is: how do we run such a tool in a new home, without asking the users to write new annotations? Consider a social researcher who is interested in deploying our tool with couples to study their routines and interactions. How would the researcher create a classifier for each home that identifies the occupants of that particular home? Does the researcher ask the occupants of the house to look at a set of trajectories, try to remember who did what, and label the trajectories? This is clearly cumbersome and prohibitive. However, in the absence of labeled data, one cannot train a classifier to identify people in that home.

We introduce a new solution that customizes the tool for new homes without asking users to label any data. Our idea is simple: the researcher can ask the occupants of the house to wear an accelerometer for only a few days. We correlate the acceleration data with the trajectories obtained by our tool to identify the occupants, and automatically label their trajectories based on the acceleration data. We use these labels to train a CNN identity classifier for this home, as described above. This approach allows for scalable deployment in many homes at a minimal overhead for both the researchers and the home occupants. Once the CNN classifier is trained, the occupants need not wear the accelerometer any longer, and the system operates purely on RF signals.

We implemented our design as a stand-alone home sensor called Marko. Once the sensor is connected to the AC power and the Internet, it can passively collect data for months without any additional user overhead. We deployed this sensor in 6 homes for a period of one month. The homes have 2 to 4 residents, whose age varies between 21 and 84 years. Each home is occupied by its actual residents who go about their lives, without any restrictions on their movements or interactions. We use these deployments to evaluate our algorithms and demonstrate the accuracy of Marko (Section 6). We illustrate the benefits of our system with case studies. They show how Marko tracks the interaction between a patient and his caregiver, provides a doctor with useful information to adjust medications, and reveals the routine of a couple and their interaction with each other.

Contributions: This paper makes the following contributions.

- It introduces the first RF-based passive system that enables in-home user identification and behavioral sensing.
- It presents an algorithm for identifying users in their homes using RF reflections, without any restrictions on their movements or the presence of multiple people.
- It develops an automatic labeling approach to bootstrap the system in new homes without additional human labeling efforts.
- It presents three case studies demonstrating the potential for performing passive studies of caregiver-patient interaction, couple's routine at home, and functional profiling of patients.

Our system makes the first step toward low-overhead and passive behavioral sensing in homes using radio reflections. We believe it can serve as a building block for new HCI capabilities to enable smart environments, non-invasive health sensing, and understanding user interactions.

2 RELATED WORK

Self-reporting through diaries and questionnaires is the most common way to measure people's behavior in homes. Self-reporting, however, can often be biased and inaccurate as people are forgetful [39, 60]. Further, the overhead of self-reporting causes subjects to stop reporting in long-term studies [39, 56]. Researchers have also leveraged smartphones and wearable sensors for behavioral sensing [6, 12, 32, 46, 56]. These systems use smartphones or wearable devices to collect various sensor data (acceleration, audio, light, etc.) and classify basic activities (e.g., walking, conversation, sleeping). While they simplify data collection, past studies have shown that wearable devices lead to adherence problems because people stop using the sensors with time [10, 16, 33, 34, 51]. In contrast, Marko uses radio reflections to sense users' locations and identities, without requiring users to wear sensors, write diaries, or answer questionnaires.

Recent advances in passive RF localization [4, 27, 37] provides new opportunities for in-home behavioral sensing. These systems transmit a wireless signal and analyze its reflections to localize nearby users, without any sensor on their bodies [4, 25, 27, 37]. These systems, however, have no notion of identity. Marko augments passive RF sensing with identification. Further, through actual in-home monitoring of real users, it demonstrates the potential of RF-based passive in-home behavioral studies.

The problem of passive RF-based identification has received attention in recent years motivated by the increased applications of passive RF sensing [5, 24, 25, 45, 47, 58, 59]. However, past work on RF-based identification is restricted to environments with a single user [3, 9, 23, 50, 57, 62, 64] Further, most prior solutions restrict the user to walking on one or a few predetermined paths without stopping or turning [3, 9, 23, 50, 57, 62]. Even small deviations from the predetermined path (e.g., shifting the walk by 1 meter) can cause major changes in the signal and a significant reduction in identification accuracy [9, 57]. We aim to address those challenges to allow for identifying users in their own homes, while they live their normal lives, and in the presence of multiple people in the environment. Further, we develop a practical solution to automatically label RF data with user identities for new homes and new users

There is also research on user identification based on facial recognition, gait, floor or ceiling sensors, ultrasonic and audio sensors, wearable sensors and on-body RFID tags [7, 18, 20, 22, 28, 30, 35, 36, 38, 41, 42, 48, 49, 63]. In contrast, this paper enables identification purely base on RF signals. Passive RF systems are less intrusive than cameras and audio systems, making them more suitable for homes and private spaces. They are also easier to deploy in comparison to instrumenting the entire floor or ceiling with sensors. They also do not require users to wear or charge any sensors. Further, the paper integrate identification within a tool for passive in-home behavior sensing and demonstrate the benefit of the design using three case studies.

3 MARKO'S SYSTEM DESIGN

Marko is a wireless sensor that passively collects information about users' behavior in the home. Marko transmits an RF signal and processes its reflections into short clips of RF videos (RF frames) and short user trajectories (tracklets). Marko has a convolutional neural network (CNN) that tags each tracklet with the corresponding user identity. Given the home floor plan, Marko shows how users interact with space at home and each other. For example, for a couple living together, Marko provides enough information to answer questions of the form: Does the couple sleep in the same room/bed? Who wakes up first? Who prepares dinner? Does the couple eat together? etc.

Next, we describe how Marko processes the RF-signal to extract RF frames and tracklets, and how it identifies people. Later in the evaluation section, we describe three case studies where Marko reveals various behavioral patterns.

Processing RF Signals

Marko processes the RF signals from the radio to extract two types of data: RF frames and tracklets.

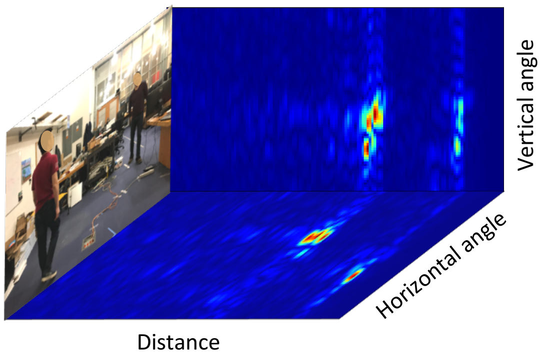
(a) RF frames: As in past passive RF-based localization systems [3, 4], Marko uses an FMCW radio equipped with two antenna arrays: a horizontal array and a vertical array. This combination allows for separating RF signals that arrive from different locations in space. Specifically, we can compute RF reflections from different locations in the horizontal plane:

$$P(d, \theta) = \sum_{n=1}^{N} \sum_{t=1}^{T} s_{n,t} e^{j2\pi \frac{kd}{c} t} e^{j2\pi \frac{nl\cos\theta}{\lambda}}, \qquad (1)$$

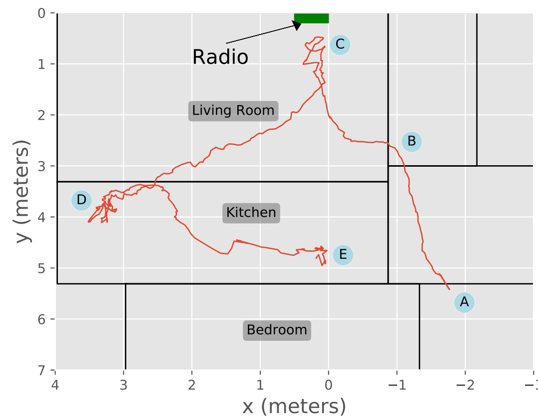
where $P(d, \theta)$ is the signal from distance d and azimuthal angle θ, N is the number of antennas in the horizontal array, T is the number of samples in an FMCW chirp, $s_{n,t}$ is the signal received by antenna n at the t^{th} sample in the chirp with slope k, l is the antennas spacing, λ is the wavelength, and c is the speed of light. Similarly, we can project the received signal on a vertical plane by substituting the signal from the vertical array in the above equation.

Thus, at each time step, we can represent the RF signal using its projection on two planes: a horizontal plane and a vertical plane. Figure 1a shows an example of the RF signal captured by the two arrays. The RF signal is a complex number, hence, we plot its magnitude in the figure. The figure shows that the horizontal plane separates people based on their locations whereas the vertical plane captures information about their height and build. We refer to these planes as the vertical and horizontal RF frames.

Marko operates on sequences of such horizontal and vertical RF frames. By processing sequences of RF frames, Marko can capture both spatial features related to people's heights and body shapes, and temporal features related to their gait



(a) Horizontal and vertical RF frames at a given time. Red and blue pixels refers to large and small values. The frames show high values at spatial locations that correspond to the two users.



(b) One tracklet (red) processed from the RF signals. The green rectangle on the top is Marko's radio. The black solid lines are room boundaries.

Figure 1: Visualization of the data Marko operates on: RF frames (left) and tracklet (right).

and the way they move. Note that in contrast to past work that uses WiFi CSI or Doppler effect, the RF frames have spatial information, providing more explicit information about user identities.

(b) Tracklets: Marko leverages past work on wireless localization to extract user location from RF signals [4, 27]. The basic idea is simple, each person appears as a blob in the horizontal frame, and the location of the user can be estimated as the center of that blob. (For the detailed localization procedure, please see [4]).

Once we have user locations, we connect location measurements in consecutive frames to create short trajectories which we call tracklets. We initialize a new tracklet whenever there is a big jump in the location (larger than 50 cm). Otherwise, each tracklet is extended with the closest location in the next time frame. (Our radio generates 30 frames per second, hence one can assume continuity of locations.) We apply a Kalman filter on the location measurements for all tracklets to handle noisy measurements.

Figure 1b shows an example tracklet along with the floor plan of the home. The person walks into the coverage area of Marko at (A), and makes a short turn at (B) to get into the living room. He then goes to the cabinet (C), probably picks up something, and turns to the dinning table at (D), stays for a while, and finally leaves the device's coverage area at (E).

The tracklet in Figure 1b shows that users do not walk on straight lines and tend to make unpredictable moves. Also, it is worth noting that identification has to be performed repeatedly every time the user exits the coverage area or when the device loses the person due to occlusion by metallic obstacles (e.g., mirrors and TV screens). For example, in Figure 1b, most of the bedroom is outside the coverage area of the radio, and hence the user has to be re-identified every time he/she goes to the bedroom and comes back.

Identification with RF Signals

The identification module is based on a convolutional neural network classifier. It operates over windows of 5 seconds[1]. For each window, given the RF frames and the corresponding tracklets from the same time interval, the identification module tags each tracklet in the input window with an identity.

(a) Removing extraneous information: The RF signal contains a lot of extraneous information that is not useful for identification. Thus, we first eliminate RF reflections from static objects (e.g., furniture) by subtracting consecutive RF frames. This eliminates all static objects whose reflections do not change over time and leaves signals from moving people.

Still, the RF signal collected from homes corresponds to many activities and situations. Different activities performed by the same person may result in much larger differences in the RF signal than differences in the identity of the person performing the activity. Asking the CNN to learn all possible activities and situations is unlikely to work. In that case, the neural network tends to learn the noise caused by different human activities and ignore features that actually distinguish different people. This leads to over-fitting the training data and poor generalization on test data.

Thus, instead of feeding the neural network all the data, we focus on walking periods –i.e., periods during which the user changes location. Of course the person can walk on a curved trajectory, make turns, slow down, etc. To identify walking periods, for each time window, we estimate the diameter of a

[1]We choose 5 seconds since it is long enough to capture user gait but also short enough to allow for quick identification.

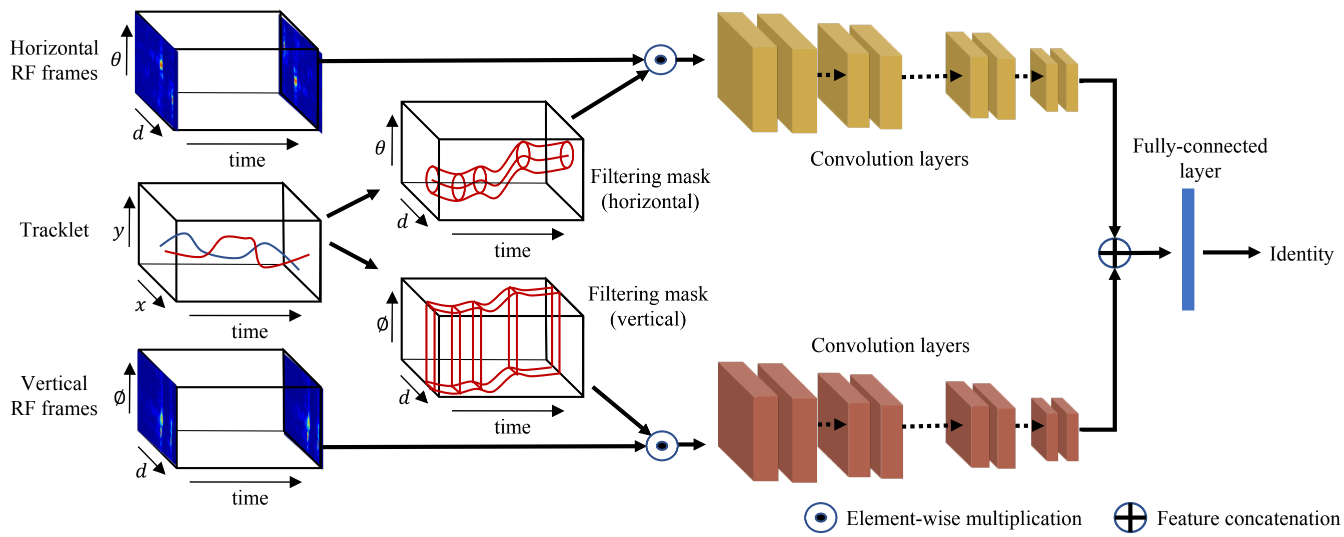


Figure 2: The identification CNN. The network takes as input a sequence of RF frames and the tracklets from the same time period. It tags each tracklet by the predicted user identity. The network has spatial filtering masks that allow it to focus on one person at a time. It also has two branches for processing information from the horizontal and vertical antenna arrays.

circle that bounds the tracklet using the diameter estimation algorithm in [25]. If the diameter of the tracklet exceeds a distance threshold (1 meter), we say the user is walking and pass the RF frames and tracklet to the CNN. As the user walks, the CNN estimates her/his identity over windows of 5-second wide. If the user stops walking and stays in the same location, the identity is assumed to persist.

(b) Dealing with multiple users: We need to deal with scenarios with multiple users. It is quite often that multiple users can be moving in the same time window. Directly feeding such data to the CNN is confusing, since it is not clear which user's identity the network should learn or predict. To address this problem, we leverage the tracklet to spatially separate the RF reflections from each user. We can then operate on each user's reflections separately and infer his/her identity.

Specifically, to focus on the RF signals of a user for a given tracklet, for each time step, we take a circle of radius r around her location in the horizontal frame[2]. For the vertical frame, we assume that the signals reflected from different elevations at that circle are from the same person. This process creates two filtering masks, a horizontal and a vertical mask, that allow us to focus on the signal from one person at any time. Below, we reinvestigate these masks and describe how they interact with the CNN.

(c) CNN architecture: For each home, we train a CNN for identifying the users. Figure 2 shows a schematic of our network. The network has two branches, one for horizontal frames and another for vertical frames. The two branches are combined in a final layer that aggregates their information.

Figure 2 also shows that our CNN takes care of separating RF information from different users so that when multiple people move together each of them is identified accurately. Specifically, the figure shows an example of two users moving at the same time, which is expressed using two tracklets: red and blue. The example in the figure focuses on the person moving along the red tracklet. The CNN uses the horizontal and vertical masks described above, which are illustrated by the red 3D tubes in the figure. The CNN takes the dot product of the horizontal and vertical frames with their corresponding masks. This has the effect of keeping only the RF signals reflected from the person moving along that red tracklet, while zeroing out all other RF signals. The blue tracklet is similarly processed using its corresponding masks.

The CNN architecture also takes care of capturing both spatial features related to a person's height and build as well as temporal features related to movement dynamics and gait. Specifically, each layer in our network uses spatio-temporal convolutions to aggregate information across space and time. Each branch in the CNN has 10 layers with a kernel of 5×3×3 on the three dimensions[3]. Following design practices in visual recognition [52], we double the number of channels and halve the dimensions every other layer. We perform average pooling at the last layer in both the spatial and temporal dimensions to create the feature vector. The output

[2]Our default r is 50 cm.

[3]The number of layers in a CNN has to be large enough to abstract the information and small enough to avoid over-fitting. The number of layers and kernel sizes are chosen empirically based on our data.

(a) Acceleration sensor **(b) Sensor with an ankle strap**

Figure 3: The acceleration sensor provides ground truth user identity for training the identification CNN. The user wears the sensor only for a few days to collect training data. During the operation mode, Marko works without any wearables.

features of the two CNNs are concatenated before the final fully-connected layer, which predicts the identity.

The CNN is trained by minimizing the cross-entropy loss:

$$\min_{\theta} \sum_{n=1}^{N} \sum_{m=1}^{M} -y_{m,n} \log \hat{p_m}(x_n; \theta), \tag{2}$$

where N is the total number of 5-second windows, M is the number of users to classify, $y_{m,n}$ is the binary indicator if label m is correct for example x_n, and $\hat{p_m}(x_n; \theta)$ is the predicted probability that example x_n is person m given model parameters θ.

During testing, the person m^* with the highest predicted probability given the RF frames x_i is used to tag the corresponding tracklet:

$$m^* = \operatorname*{argmax}_m \hat{p_m}(x_i; \theta). \tag{3}$$

4 AUTOMATIC LABELING FOR NEW HOMES

Typical studies may deploy the device in tens or hundreds of homes. Every home has different occupants, and hence needs a new CNN classifier. Training such a classifier requires labeled data from each home. Asking the subjects in each home to label a few days of their data for training creates too much overhead. It is hard for people to look at trajectories and try to remember who did what and when. We need a solution that automatically create labeled data for deployments in new homes.

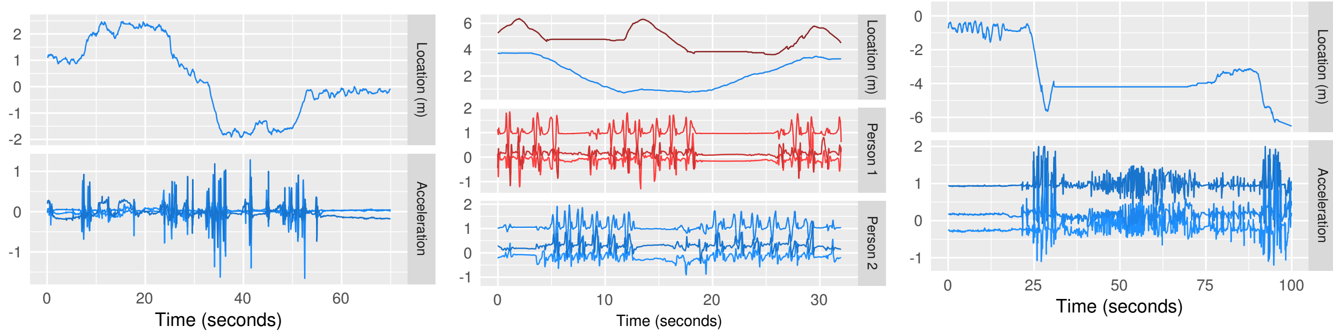
Next, we present a solution that automatically generates labeled data for deployments in new homes. When Marko is deployed in a new home, users interested in being identified by the system wear a small accelerometer for a few days. During that period, Marko collects RF signals and acceleration data. It processes the RF signals to extract RF frames and user tracklets (Section 3). For each tracklet, Marko correlates the motion along the tracklet with the acceleration from the wearable sensors; it labels the tracklet with the identity of the user whose acceleration matches the motion in the tracklet. Once the system has enough labeled data, the users can stop wearing the accelerometer, and identification is subsequently performed using RF signals alone. Note that wearing an accelerometer for a few days to train the system is significantly less onerous than requiring the user to wear and charge sensors for an indefinite period of time.

We use a small accelerometer [1] shown in Figure 3a. We recommend that the user wears the sensor using an ankle strap as in Figure 3b. The sensor streams acceleration data to the Marko device using the Bluetooth Low Energy protocol. The battery lasts for around one week, longer than the training phase. Thus, the user does not need to remember to charge the acceleration sensor.

(a) Acceleration Data: Figure 4 shows multiple examples of tracklets and acceleration data recorded at the same time. To simplify the visualization of a tracklet, we plot how the distance from the device changes over time. In Figure 4a, the top graph shows the change in distance for a particular tracklet, and the bottom graph shows the corresponding 3-axis acceleration data. It is clear that when the user walks, the acceleration oscillates with the steps. The acceleration becomes a flat line once the person stops moving. However, one should be careful when matching acceleration with motion as multiple users may be moving at the same time as in Figure 4b. Also, the acceleration can change when a user is stationary because he is tapping with his feet or moving his body in place, as in Figure 4c. Hence, we need a robust algorithm for matching acceleration with tracklets.

(b) Labeling network: We formulate the labeling problem as measuring the similarity between data from the two modalities: acceleration and tracklets. We cannot simply correlate the acceleration with the tracklet to compute their similarity because they are quite different. Instead, we design a neural network that takes a tracklet and the acceleration data from the same time period, and learns to produce a similarity score. For segments of data with high similarity scores, we assign the corresponding user identity to the tracklet.

Model design: The architecture of the labeling network is shown in Figure 5. The neural network takes an entire tracklet and acceleration data of the same period as input. For shorter tracklets, we pad zeros to make them the same length. The network has two branches where one branch is devoted to modeling acceleration and the other models tracklets. Both branches of the network use 3 layers of convolutions along the temporal dimension with a kernel size of 3. Each convolutional layer is followed by a ReLU activation and a dropout layer. At the end of the two branches, we have two feature vectors representing the two data types. We use the dot product of the feature vectors to represent their similarities at each time step. Finally, we perform max pooling



(a) Single person walking. **(b) Two people (red and blue) walking.** **(c) A complicated example of acceleration and location data of a same person.**

Figure 4: Passive RF location data and wearable acceleration data. The top row in each figure shows the location in one dimension, and the bottom row shows the 3-axis acceleration data.

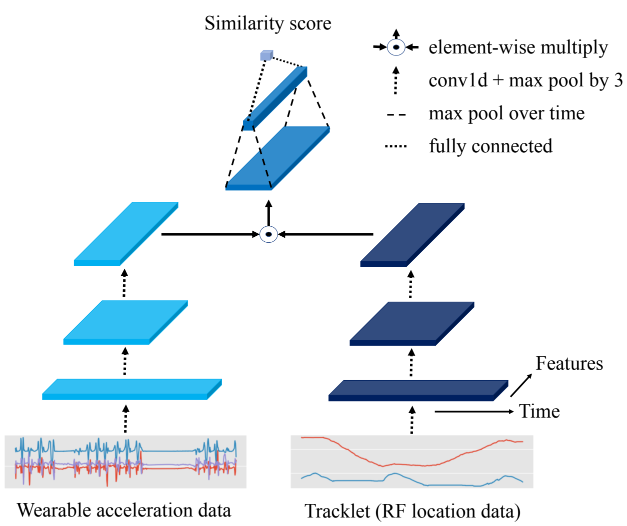


Figure 5: The labeling network. A two-branch neural network that learns a similarity measure between wearable acceleration data and tracklets. If the two data streams are similar, the tracklet is labeled with the identity of the user carrying the acceleration sensor.

in the temporal dimension and use a fully connected layer to produce a single similarity score for the entire period.

Training: To train the network, we collect acceleration and tracklets generated by the same person as correct examples. We also randomly assign acceleration streams to tracklets to create wrong examples. In training, we set the ground truth similarity scores for the correct pairs as 1 and wrong pairs as 0. We train the network by minimizing the cross-entropy loss between the true and predicted similarity scores using stochastic gradient descent with Adam optimizer [31].

Finally, we note that once the labeling network is trained, it can label data automatically from any new home – i.e., there is no need to train a labeling network per home. This is in contrast with the identification network – since each home has different occupants with their own identities, the identification CNN has to be trained per home. The nice thing, however, is training the identification CNN can be automated using one labeling network.

5 IMPLEMENTATION

We built the FMCW radio using off-the-shelf components based on the design of [4]. The radio repeatedly generates a frequency chirp ranging from 5.46 to 7.24 GHz[4]. The two antenna arrays are integrated with the radio on a printed circuit board (PCB)[5]. A single board computer processes the RF signals from the PCB, and sends data to the cloud over WiFi. In the training phase, it also handles the connection with the acceleration sensor using Bluetooth Low Energy. All the components are packed in a stand-alone box with a size of $12 \times 15 \times 1.5$ inches, and can be easily deployed by hanging it on the wall.

6 EVALUATION

We evaluate Marko through deployments in 6 homes for a period of one month[6]. Our deployments are with users whose age ranges between 21 and 84. One of the homes is a townhouse with 4 residents. Three of the homes are one-bedroom apartments each hosting a young couple with no children. The other two homes are in an assisted living facility, where the residents live in their own unit and receives frequent visits from the nurses.

During the deployment, the residents go about their normal lives with no restrictions. They wear different clothes every day, are free to stay anywhere inside the home, and can move furniture or chairs just as they normally do.

[4]The average power of the radio complies with FCC regulations [11]
[5]While we build our own radio for flexibility, various FMCW radios with antenna arrays are also available on the market [2, 61]
[6]The deployments are approved by our institutional review board (IRB).

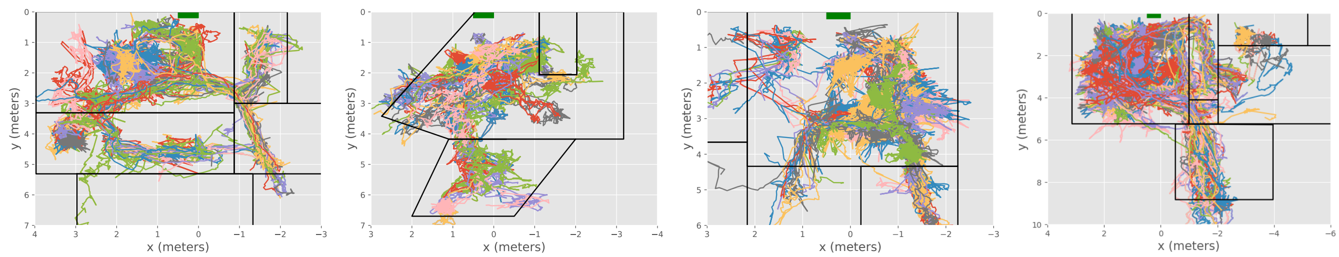


Figure 6: A representative set of tracklets collected by Marko in different homes. The green rectangle box at the top is the location of Marko's radio. The black lines are boundaries of different rooms or areas. Different tracklets are plotted in different colors.

Data Collected in the Wild

Before evaluating Marko, we show a representative set of the tracklets collected in different homes to give the reader some insight into the underlying motion patterns. Figure 6 shows one day of tracklets from four homes. The figure shows that people's trajectories can be quite complex and far from straight lines.

Evaluation of Automatic Labeling

We start by evaluating the automatic labeling module which matches acceleration data with tracklets. For each home, we randomly sample a person and ask the person to wear the accelerometer. For the homes in the assisted living facility, we ask the residents to wear the accelerometer. We use that data to train the matching network.

To obtain the ground-truth labels for matching tracklets and acceleration, our first attempt was to ask the user to write a diary of her movements at room level. Unfortunately, the diaries are incomplete, sometimes missing a full day or multiple hours in a day. Thus, we ask three human labelers to manually label the ground truth for each home, while using the information in the diaries for guidance. We implement a labeling tool that visualizes both tracklets and acceleration data as in Figure 4a. The human labelers are asked to label tracklets whose motion matches the acceleration data. The three labelers compare their labels and resolve disagreements. When in doubt, the labelers call the subjects to ensure the labeling is correct and discard examples with disagreements that cannot be resolved.

Accuracy of Automatic Labeling: We evaluate the accuracy of our automatic labeling network and compare it to the ground truth obtained by human-labeling as described above. Recall that the main point of automatic labeling is to label data in new homes not seen in training. Thus, we perform leave-one-out evaluation–i.e., for each test home, we train a model excluding the test home's data.

Table 1 shows the test accuracy of the automatic labeling network in each home. The results show that our model has an average accuracy of 95% when labeling data in a new home that it has not trained on. The accuracy is close to the inter-labeler agreement rate (which is 96%). In fact, the misclassified examples are typically those the human labelers have trouble labeling.

Environments	Accuracy
Home 1	97%
Home 2	96%
Home 3	95%
Home 4	94%
Home 5	94%
Home 6	94%

Table 1: Accuracy of automatic labeling. The table shows that the labeling network has high accuracy and works accurately even when it is trained and tested on different homes.

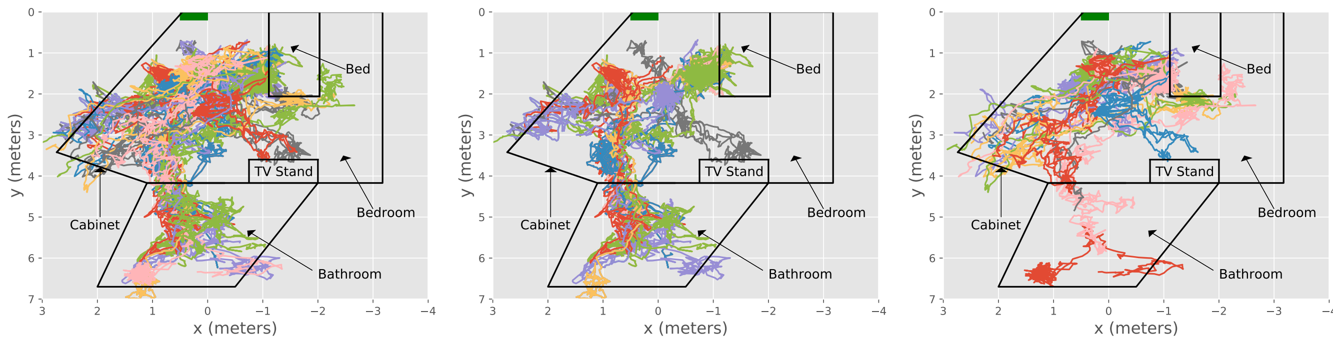
Evaluation of RF-Based Identification

In this section, we evaluate the identification CNN and shed light on how it works.

(a) Identification Accuracy: We present the identification results from our deployments in the homes. For each home, we train a CNN classifier using the data from that place. We use 80% of the data for training and 20% for testing. Since the objective of the classifier is to predict the future, we use earlier data for training and later data for testing. All of the reported results are test accuracy.

We evaluate our model based on average accuracy where the average is taken across all classes. We also compare Marko with a baseline that takes the same input including the horizontal and vertical masks and uses a random forest classifier [8] as opposed to our spatio-temporal CNN. For the baseline classifier to achieve better results, we had to reduce the input dimensions by applying Principal Component Analysis (PCA) on the RF frames and keeping the principal components that explain 99% of the variance. We experiment with different baseline parameters (e.g., tree depth, number of samples in a leaf node, etc.) and report the best results.

Table 2 summarizes the results. Our identification accuracy ranges between 85% and 95%, and the average across all homes is 90%. The difference in accuracy between homes are due to the intrinsic differences between the occupants.



(a) All tracklets (without tagging identities) **(b) The patient's tracklets** **(c) The nurse's tracklets**

Figure 7: One day of tracklets. The figure compares all tracklets of the patient and those of the nurse. The green rectangle on top is the location of Marko's radio. The black lines refer to the boundaries between rooms, and the marking of the bed area.

Environments	Marko	Baseline
Home 1	95%	66%
Home 2	88%	58%
Home 3	92%	57%
Home 4	92%	77%
Home 5	89%	54%
Home 6	85%	59%

Table 2: Identification Accuracy. The table shows that Marko achieves high accuracy and significantly outperforms a random-forest baseline that is trained and tested using the same data.

For example, we notice older couples typically have larger differences between their health conditions than younger ones, which leads to more distinct movement patterns. In all cases, however, the accuracy of Marko is much higher than the baseline. This shows that our spatio-temporal CNN is better at extracting the complex patterns in RF reflections that are relevant to identification.

(b) Accuracy with Multiple Moving Users: Most of the time, the homes in our study have multiple people, and hence the above results cover multiple users. Here, however, we focus on the harder case when multiple users in the home are walking at the same time. Recall that in such cases, we use the tracklets to create spatial filtering masks that allow the CNN to focus on one user at any time. We then repeat the process for each user (see Section 3). Table 3 shows the identification accuracies with and without our filtering masks. Without the filtering masks, the accuracy is basically random. This is because with multiple moving users in the scene, the RF signal (without the masks) has information about more than one identity, and hence the CNN can be easily confused.

Users' Behaviors and Interactions in the Homes

In this section, we demonstrate how Marko enables new ways of behavioral sensing in homes. We look at three case studies using data from our deployments. Two studies are in

Scenarios	Accuracy
Multiple users (with filtering masks)	81%
Multiple users (without filtering masks)	50%

Table 3: The importance of spatial filtering masks. The table shows that the filtering masks are necessary to achieve accurate identification in the presence of multiple users.

the health care domain, where we are interested in caregiver-patient interaction and functional profiling of a patient. One study is related to social interactions, where we study how a couple spent time at home and their interaction.

(a) Case Study I: Caregiver-Patient Interaction: In one of the homes within the assisted living facility, we monitor the elderly patient and nurses who take care of him. Both his family and medical doctors of the facility are interested in ensuring that the patient receives proper care from the nurses. Also, they are interested in the overall health condition of the patient. We use our data to answer some of their questions.

We use Marko to distinguish tracklets of the patient from those of the nurses. Figure 7a plots one day of tracklets, and Figure 7b and Figure 7c show the differences between the patient's tracklets and those of the nurse. The green filled rectangle on top of each figure is the location of the Marko device. The black lines refer to the boundaries between rooms, and the marking of the bed area.

Interestingly, once we separate the tracklets of the patient from those of the nurse, we discover distinct behavioral patterns. For example, we discover that the elderly resident never goes to the lower left corner of his bedroom, which is marked as the cabinet in Figure 7b and Figure 7c. After we talked to the doctor at the facility, we learned that the medication cabinet is locked in a closet at that corner. Only the nurse is allowed to give medication to the patient. By tracking when the nurse visits that corner and then attends to the patient, we can infer if the patient took his medications at the prescribed times. Similarly by tracking identity, we

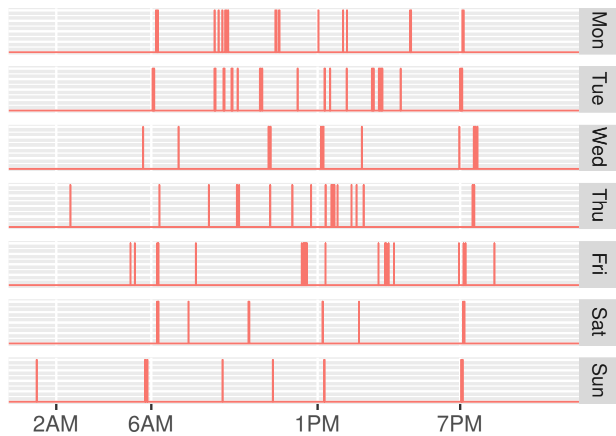


Figure 8: Times of nurse visits in a week.

can tell when the patient goes to bed and distinguish that from the nurse making the bed. We can also see tracklets going to the TV stand and infer whether the patient usually turns the TV on/off on his own or the nurse did it for him.

We can also analyze the visit patterns of the nurse. We look at one week of data and plot the visit times for each day in Figure 8. The figure shows that every day the nurse visits the resident around 6 am and 7 pm. Those are the times when the nurse wakes the patient up in the morning and helps him get to bed in the evening. She also administers medications at those times. The nurse also administers medication at 1pm. The rest of the visit are less regular and spread throughout the day. On some the days, there are more visits than usual. It is because the frequency of the nurse's visits typically depends on the resident's health condition.

For example, we notice that the nurse visited on Thursday and Sunday night around 3 am and 1 am, which are the times when the resident is typically asleep. After checking with the facility, we learned that on those days the resident was wandering at night. The nurse came and helped him get back to his room. By analyzing the tracklets tagged with identities, we can estimate the level of care a patient receives.

(b) Case Study II: Functional Profiling of a Patient: The resident in the second home in the assisted living facility is a patient who suffers from dementia and severe agitation. To adjust her agitation medications, the doctor typically asks the nurses how agitated she seems. However, he has no way to objectively quantify her agitation.

We use Marko to monitor the patient behavior and look for signs of agitation. A key symptom of agitation is pacing [53]. Thus, we separate tracklets of the patient from those of the nurses. We consider tracklets that show a pacing behavior where the patient repeatedly moves back an forth between two locations. We discover that the patient repeatedly goes from the door to the bed and back, and keeps moving back and forth along this path more than 100 times per day. We

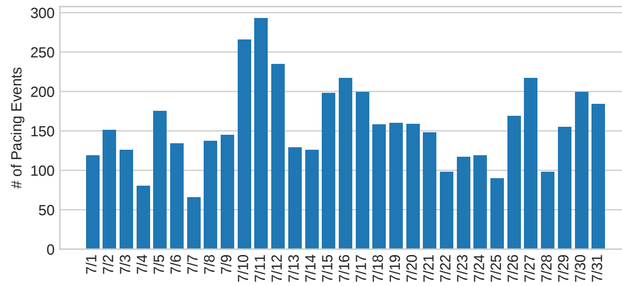


Figure 9: Number of pacing events across a month.

plot the number of such pacing events per day in Figure 9. The figure shows that the patient paces excessively, which reveals a high level of agitation. Further, occasionally, there is a surge in her pacing behavior that lasts for 2 to 3 days. After matching her pacing behavior with her medical record, the doctor discovered that she paces more on the days following a visit from her family. An example of such surge happens on on July 10th. This shows that Marko provides a new passive way for functional health profiling of patients at home.

(c) Case Study III: Couple's Routines and Interaction: In this study, we use Marko' data to analyze the daily behavior of a couple at home. We pick one representative day of a particular couple from our deployments. We look through the identified tracklets and plot how the man and woman spend time at home in Figure 10. The figure shows that the couple has very different schedules. Both of them get up early around 8 am as they leave the bedroom. The woman has regular work hours as she always leaves around 9 am and comes back at 6 pm. After coming home, she makes her dinner in the kitchen and spends time in the living room. On the other hand, the man has longer work hours and rarely has dinner at home. He leaves at 8 am and comes back at 9 pm. Despite long work hours, the couple still spend time together in the living room before going to sleep. Interestingly, the woman wakes up after mid-night and stays in the living room for a few hours. We later found that the woman could not fall asleep on that night. She was watching TV and eventually fell asleep on the sofa.

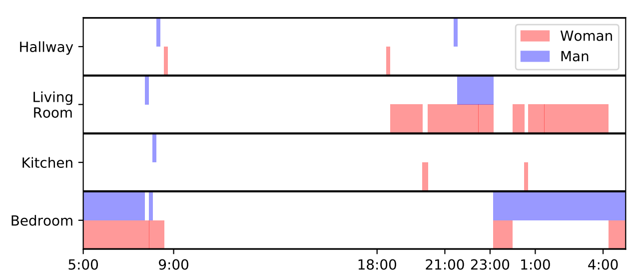


Figure 10: A timeline for a day of a couple. It shows how the couple has different daily schedules, different ways to spend time at home, and when and where they interact.

Additional Analyses and Evaluation

We present additional analyses of how the labeling and identification network operate, and evaluate our design choices in the supplementary materials.

7 CONCLUSION AND DISCUSSION

This paper introduces Marko, the first system that uses passive RF reflections to enable identification and behavioral sensing in homes. It also provides a solution to customizing the system to new homes without asking users to annotate any data. We evaluate Marko through real-world deployments in 6 homes over a period of one month, and demonstrate its value for studying couple relationships and caregiver-patient interaction.

We would also like to note the importance of ensuring the technology does not get misused to infringe on privacy. Research in this domain must follow IRB regulations and abide by users' signed consent that specifies data access and storage policies. General policies about the use of personal data similar to those taken by Europe are also helpful [43].

Overall, we believe this work takes an important step towards low-overhead and passive behavioral sensing in homes. As such it enables new HCI capabilities for smart environments, continuous health sensing, and understanding users' routines and interactions.

REFERENCES

[1] 2018. mbientlab MetaMotion sensors. Retrieved September 20, 2018 from https://mbientlab.com

[2] 2018. Walabot. Retrieved September 20, 2018 from https://walabot.com/

[3] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 219.

[4] Fadel Adib, Zachary Kabelac, and Dina Katabi. 2015. Multi-person localization via RF body reflections. In *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)*. 279–292.

[5] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 837–846.

[6] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. The social fMRI: measuring, understanding, and designing social mechanisms in the real world. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 445–454.

[7] Chuankai An, Tianxing Li, Zhao Tian, Andrew T Campbell, and Xia Zhou. 2015. Visible light knows who you are. In *Proceedings of the 2nd International Workshop on Visible Light Communications Systems*. ACM, 39–44.

[8] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[9] Yuanying Chen, Wei Dong, Yi Gao, Xue Liu, and Tao Gu. 2017. Rapid: A Multimodal and Device-free Approach Using Noise Estimation for Robust Person Identification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 41.

[10] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 647–658. https://doi.org/10.1145/2750858.2807554

[11] Federal Communications Commission. 2018. Rules and Regulations for Title 47.

[12] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1797–1806.

[13] Ann C Crouter, Maureen Perry-Jenkins, Ted L Huston, and Duane W Crawford. 1989. The influence of work-induced psychological states on behavior at home. *Basic and Applied Social Psychology* 10, 3 (1989), 273–292.

[14] Daniela Fialová, Eva Topinková, Giovanni Gambassi, Harriet Finne-Soveri, Pálmi V Jónsson, Iain Carpenter, Marianne Schroll, Graziano Onder, Liv Wergeland Sørbye, Cordula Wagner, et al. 2005. Potentially inappropriate medication use among elderly home care patients in Europe. *Jama* 293, 11 (2005), 1348–1358.

[15] Gerda G Fillenbaum. 1985. Screening the elderly: A brief instrumental activities of daily living measure. *Journal of the American Geriatrics Society* 33, 10 (1985), 698–706.

[16] Glenn Fox, Shaun Garland, Andrew Keibel, and Leslie Saxon. 2017. Why People Stick With or Abandon Wearable Devices. https://catalyst.nejm.org/stay-abandon-wearable-devices/

[17] Hulya Gokalp and Malcolm Clarke. 2013. Monitoring activities of daily living of the elderly and the potential for its use in telecare and telehealth: a review. *TELEMEDICINE and e-HEALTH* 19, 12 (2013), 910–923.

[18] Tobias Grosse-Puppendahl, Xavier Dellangnol, Christian Hatzfeld, Biying Fu, Mario Kupnik, Arjan Kuijper, Matthias R Hastall, James Scott, and Marco Gruteser. 2016. Platypus: Indoor Localization and Identification Through Sensing of Electric Potential Changes in Human Bodies. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 17–30.

[19] Abdelsalam Helal, Diane J Cook, and Mark Schmalz. 2009. Smart home-based health platform for behavioral monitoring and alteration of diabetes patients. *Journal of diabetes science and technology* 3, 1 (2009), 141–148.

[20] Timothy W Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. 2012. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 309–322.

[21] Sandra L Hofferth and John F Sandberg. 2001. How American children spend their time. *Journal of Marriage and Family* 63, 2 (2001), 295–308.

[22] Christian Holz, Senaka Buthpitiya, and Marius Knaust. 2015. Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 3011–3014.

[23] Feng Hong, Xiang Wang, Yanni Yang, Yuan Zong, Yuliang Zhang, and Zhongwen Guo. 2016. WFID: Passive Device-free Human Identification Using WiFi Signal. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 47–56.

[24] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. 2017. Zero-Effort In-Home Sleep and Insomnia Monitoring using Radio Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 59.

[25] Chen-Yu Hsu, Yuchen Liu, Zachary Kabelac, Rumen Hristov, Dina Katabi, and Christine Liu. 2017. Extracting Gait Velocity and Stride Length from Surrounding Radio Signals. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2116–2126.

[26] Remus Ilies, Kelly M Schwind, David T Wagner, Michael D Johnson, D Scott DeRue, and Daniel R Ilgen. 2007. When can employees have a family life? The effects of daily workload and affect on work-family conflict and social behaviors at home. *Journal of Applied Psychology* 92, 5 (2007), 1368.

[27] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. 2015. WiDeo: fine-grained device-free motion tracing using RF backscatter. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation*. USENIX Association, 189–204.

[28] Avinash Kalyanaraman, Dezhi Hong, Elahe Soltanaghaei, and Kamin Whitehouse. 2017. Forma Track: Tracking People based on Body Shape. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 61.

[29] Sidney Katz. 1983. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *Journal of the American Geriatrics Society* 31, 12 (1983), 721–727.

[30] Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. 2016. Nonintrusive Occupant Identification by Sensing Body Shape and Movement. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 1–10.

[31] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[32] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications magazine* 48, 9 (2010).

[33] Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 635–646. https://doi.org/10.1145/2750858.2804288

[34] D Ledger and D McCaffrey. 2014. Inside wearables: How the science of human behavior change offers the secret to long-Term engagement. 93 (01 2014), 36–45.

[35] Lily Lee and W Eric L Grimson. 2002. Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 155–162.

[36] Hanchuan Li, Peijin Zhang, Samer Al Moubayed, Shwetak N Patel, and Alanson P Sample. 2016. Id-match: A hybrid computer vision and rfid system for recognizing individuals in groups. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4933–4944.

[37] Xiang Li, Shengjie Li, Daqing Zhang, Jie Xiong, Yasha Wang, and Hong Mei. 2016. Dynamic-music: accurate device-free indoor localization. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 196–207.

[38] Wen-Hau Liau, Chao-Lin Wu, and Li-Chen Fu. 2008. Inhabitants tracking system in a cluttered home environment via floor load sensors. *IEEE Transactions on Automation Science and Engineering* 5, 1 (2008), 10–20.

[39] Andreas Möller, Matthias Kranz, Barbara Schmid, Luis Roalter, and Stefan Diewald. 2013. Investigating self-reporting behavior in long-term studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2931–2940.

[40] Kristiina Moller, Philip C Hwang, and Birgitta Wickberg. 2008. Couple relationship and transition to parenthood: Does workload at home matter? *Journal of reproductive and infant psychology* 26, 1 (2008), 57–68.

[41] Robert J Orr and Gregory D Abowd. 2000. The smart floor: A mechanism for natural user identification and tracking. In *CHI'00 extended abstracts on Human factors in computing systems*. ACM, 275–276.

[42] O. M. Parkhi, A. Vedaldi, and A. Zisserman. 2015. Deep face recognition. In *Proceedings of the British Machine Vision Conference (BMVC)*.

[43] European Parliament and Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (2016), 1–88.

[44] Gregory S Pettit and John E Bates. 1989. Family interaction patterns and children's behavior problems from infancy to 4 years. *Developmental psychology* 25, 3 (1989), 413.

[45] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 27–38.

[46] Kiran K Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 281–290.

[47] Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury. 2015. Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 39–50.

[48] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1379–1384.

[49] Yu-Lin Shen and Chow-Shing Shin. 2009. Distributed sensing floor for an intelligent environment. *IEEE Sensors Journal* 9, 12 (2009), 1673–1678.

[50] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart User Authentication through Actuation of Daily Activities Leveraging WiFi-enabled IoT. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 5.

[51] Patrick Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John Carroll. 2015. Use and adoption challenges of wearable activity trackers. (03 2015).

[52] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[53] Kaycee M Sink, Karen F Holden, and Kristine Yaffe. 2005. Pharmacological treatment of neuropsychiatric symptoms of dementia: a review of the evidence. *Jama* 293, 5 (2005), 596–608.

[54] Robert Steele, Amanda Lo, Chris Secombe, and Yuk Kuen Wong. 2009. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. *International journal of medical informatics* 78, 12 (2009), 788–801.

[55] Andreas E Stuck, Harriet U Aronow, Andrea Steiner, Cathy A Alessi, Christophe J Büla, Marcia N Gold, Karen E Yuhas, Rosane Nisenbaum, Laurence Z Rubenstein, and John C Beck. 1995. A trial of annual in-home comprehensive geriatric assessments for elderly people living in the community. *New England Journal of Medicine* 333, 18 (1995), 1184–1189.

[56] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and

behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 3–14.

[57] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 363–373.

[58] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. ACM, 65–76.

[59] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 617–628.

[60] Rick Wash, Emilee Rader, and Chris Fennell. 2017. Can People Self-Report Security Accurately?: Agreement Between Self-Report and Behavioral Measures. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2228–2232.

[61] Luke Weston. 2014. Simple, low-cost FMCW radar. Retrieved September 20, 2018 from https://hackaday.io/project/1682-simple-low-cost-fmcw-radar#menu-description

[62] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, 4.

[63] Yang Zhang, Chouchang Jack Yang, Scott E Hudson, Chris Harrison, and Alanson Sample. 2018. Wall++: Room-Scale Interactive and Context-Aware Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 273.

[64] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7356–7365.

SUPPLEMENTARY MATERIALS

We present additional analyses and evaluation of our system to shed light on how the neural networks work and our design decisions.

Analyzing the Labeling Network

To understand how different input examples impact the similarity scores between the two data streams, we analyze how the network's output similarity score changes as it processes longer stretches of a tracklet. Figure 11 shows a scenario where the acceleration data and the tracklet are the result of movements from two *different* people. We compute the predicted score of the labeling network at time t by giving it the data from the beginning up to time t.

The figure shows that as time goes by, the model sees more data and its output similarity score also changes accordingly. In particular, in the first 10 seconds, the two data streams correlate well. The person is moving and at the same time we can see changes in the acceleration that are indicative of a person's walking. The network correctly increases its score that these two streams are from the same person. However, after the 10th second, the person stops and turns, but we still see some steps so the network decreases its score. Shortly after the 20th second, the acceleration is completely static but the person moves significantly. As a result, the similarity score drops significantly. From 40 to 60 seconds, it is clear now that the person is stationary while the acceleration keeps changing; hence, the score stays low. This shows that even if both data streams seem correlated in the beginning, the labeling network correctly predicts that the person in the tracklet is not the user wearing the accelerometer.

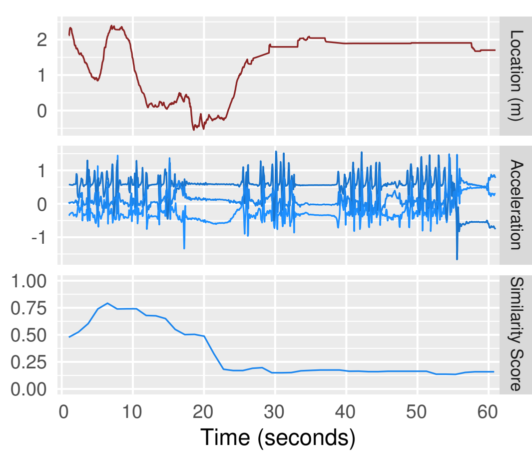


Figure 11: The output similarity score of the labeling network over time. The labeling network correctly predicts that the two streams are *not* from the same person. The output of the network adjusts correctly as it sees more data over time.

Analyzing the Identification Network

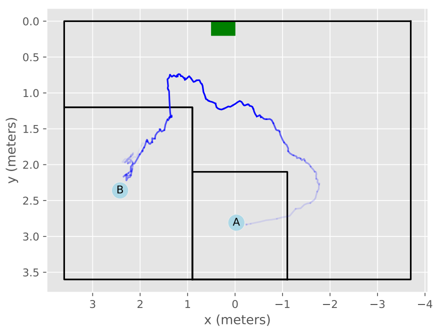
Walking vs. Stationary Periods.
We analyze the impact of training the identification CNN on walking periods. We earlier argued the importance of focusing on walking periods where the user's gait can help in differentiating users. Further, the user may be sitting or sleeping during stationary periods and hence information about her height is noisier. Table 4 compares identification accuracy as we include more stationary periods in the *training data* for Home 6. It shows that training the CNN on non-walking periods significantly degrades the performance. In particular, when the training data has 3x more stationary periods than walking periods, the testing accuracy degrades from 85% to 61%. Adding more stationary data to the training set makes it even worse, reducing the accuracy to 55%. This is because, as discussed in Section 3, different activities performed by the same user may result in much larger differences in RF reflections than differences due the user's identity. This makes it difficult for the network to glean the key features that are related to identity.

Training Data (Home 6)	Accuracy
Walking periods	85%
Walking + Stationary (3x amount of data)	61%
Walking + Stationary (13x amount of data)	55%

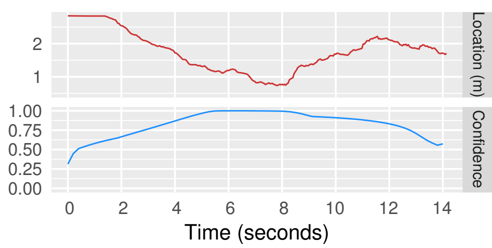
Table 4: The importance of removing stationary periods from the training data. The table shows that stationary periods have noisy information about identity. It is better to identify users as they move around and let the identities persist while they are stationary.

Model's confidence over time.
To better understand how the identification network makes predictions, we analyze the model's confidence as a person walks from one room to another. Recall that we compute the identity over windows of 5 seconds. In this experiment, we focus on a particular tracklet and the corresponding RF signal. We apply the identification model to sliding windows of 5-second each, with a step of 0.2 second. We plot the confidence scores (the predicted probability in Eq. 3) of our identification network as the window slides in Figure 12. Figure 12a shows the floor map with the studied tracklet. The tracklet starts in the bed (position A) and walks to the bathroom (position B). The confidence is shown as the intensity of the color. Figure 12b shows the confidence score and location as functions of time. For simplicity, we only show the location along the y-axis.

The figure shows that the CNN has higher confidence when the person is walking between the bed and the bathroom. In contrast, it has low confidence in the beginning (0 to 2 seconds) when the person is stepping out of bed. This is expected since when the person is still in bed it is harder for the network to identify him because it cannot extract information about gait or height. Similarly, around the end of the tracklet (12 to 14 seconds), the person makes small movements in place while in the bathroom. During that time the confidence is lower than that when he was walking. This is likely because it is easier for the CNN to identify people by modeling their gait during walking periods.



(a) Tracklet. Darker color refers to higher confidence.



(b) Prediction confidence in time.

Figure 12: Identification confidence for different points along a tracklet. While the CNN identifies the user accurately, its confidence is higher when it is looking at time windows that contain only walking and no idling or in bed motion.

Accuracy as a function of the input time window. The identification CNN takes a sequence of RF frames that span, by default, a 5-second window. Here, we analyze how the window length affects identification accuracy. Figure 13 shows the accuracy for Home 6 with different window lengths, ranging from a single frame to 6.6 seconds. As expected, a single frame leads to worse accuracy (71%) since only spatial features are captured (e.g., height and build) but no temporal information (e.g., gait) is present. As the window size increases, the accuracy improves (85% for 5-second windows), showing that temporal information in the RF signal helps with identification. Increasing the window size beyond 5-second starts degrading the performance. The reason is that much of the walking done inside the home takes less than 5 seconds. Thus, if one uses only longer periods, the training dataset becomes significantly smaller. Since deep learning naturally requires a large training set, as the training set becomes smaller, the CNN starts over-fitting the training data and cannot learn as effectively.

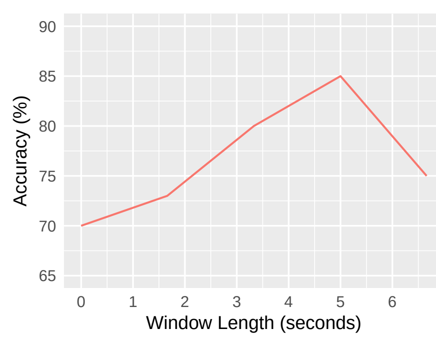


Figure 13: Identification accuracy vs. window lengths (Home 6).