Privacy-Friendly Wi-Fi-Based Occupancy Estimation with Minimal Resources

E. Makri[a,*], J. ten Brinke[b], R. Evers[b], P. Man[b], H. Olthof[b]
[a]*School of Governance, Law & Urban Development, Saxion University of Applied Sciences, The Netherlands*
[b]*School of Creative Technology, Saxion University of Applied Sciences, The Netherlands*

Abstract. Occupancy estimation is becoming an increasingly popular research topic, as solutions can be deployed both to the challenges of demand-driven ambient comfort control applications, and to the challenges of building safety and security. With our work, we aim to estimate the number of people in a particular area of a building, using only existing infrastructure. To achieve this, we collect information from the Wi-Fi Access Points installed throughout a building, in such a way that the privacy of the persons using the Wi-Fi resources remains intact. While several approaches have been proposed to address the occupancy question, our main contribution lies in that our solution uses only standard Wi-Fi infrastructure, already deployed in any modern building. In addition, we claim that our solution comes at virtually zero cost, as our mechanisms add negligible network traffic, using minimal network and processing resources, and it does not require specialised hardware.

Keywords: Occupancy number estimation, Detection, Indoors areas, Non-intrusive, Wi-Fi-based

1. Introduction

Occupancy estimation is becoming increasingly popular, especially during the last six years. It is not occupancy per se that current solutions aim to detect (i.e., answering the occupied/not occupied question), but actually estimate the number of people in a specific part of a building. This information can vastly contribute in energy savings, as HVAC (heating, ventilating, and air conditioning) systems can be automatically adjusted to the actual needs per area of a building. Since buildings account for 40% of the U.S. energy consumption (Yang et al., 2012), we can conclude that this is a very important aspect of sustainability and energy conservation.

In addition to the energy and environmental benefits that occupancy count bears, it is also of utmost importance to building safety and security. Although less emphasized in the current literature, information on the number of people in a certain part of a building, can be of great assistance to the public safety and security services. Use cases include emergency egress scenarios such as fire, and hostage situations. When emergency services (e.g., fire and police department) are aware of the number of people present in the specific area of interest, they can more efficiently manage resources. Knowing how many persons need to be rescued, and their true location within a building saves valuable resources. This is essential in emergency situations occurring at the same time, in the same region. More interestingly,

* Corresponding author
Email address: e.makri@saxion.nl (E. Makri)

proper resource management can result in timely responses, and eventually contribute in the protection of life and property. In conclusion, we argue that accurate information on the number of people in the different parts of a building is fundamental in the development of smart and safe buildings.

A lot of solutions have been proposed in the literature, with many of them achieving above 80% accuracy. However, the majority of these works (Benezeth et al., 2011; Depatla, Muralidharan, & Mostofi, 2015; Dodier et al., 2006; Dong et al., 2010; Han, Gao, & Fan, 2012; Li, Calis, & Becerik-Gerber, 2012; Meyn et al., 2009; Pan et al., 2014; Tomastik et al., 2010; Woo et al., 2011; Yang et al., 2012) requires dedicated equipment; others (Balaji et al., 2013; Benezeth et al., 2011; Christensen et al., 2014; Khan, Hossain, & Roy, 2015; Li, Calis, & Becerik-Gerber, 2012; Meyn et al., 2009; Tomastik et al., 2010) are intrusive; and some (Ebadat et al., 2013; Lam et al., 2009) assume existing sensors that are not always present in today's buildings. We aim in developing a plug and play, software-based system that is sufficiently accurate in counting the number of people in a building area, so as to serve the purposes of the aforementioned application scenarios.

We propose a Wi-Fi based occupancy estimation tool, which respects user privacy. To preserve the occupant's privacy, we deploy the current state-of-the-art hashing techniques on the collected user identifiable information, namely the IP, and MAC addresses. We consider this to be an important feature of our tool, as being able to identify where, and when a user is, can facilitate an adversary in creating user profiles that can be used from marketing and advertising purposes, to even facilitate physical crime (Stottelaar et al., 2014).

Our occupancy estimation tool is based on the Simple Network Management Protocol (SNMP). SNMP is an Internet Standard protocol for collecting and organizing information about managed devices on IP networks and for modifying that information to change device behavior. We opted for the use of SNMP, because it is standardized, and broadly used, thus making our solution ready-to-use, without requiring network modifications. The so-called SNMP traps is the built-in tool of the protocol, which allows us to collect the necessary information (i.e., usernames, MAC, and IP addresses) for our occupancy estimator. We have developed several SNMP trap filters, so as to allow fine-grained occupancy estimation throughout the buildings, where a network spans. The details of these filters' workings follow in Section 3.

Our main contributions are summarized as follows:

- **Enable non-intrusive people counting in any building with basic Wi-Fi network infrastructure.** We assume the existence of basic Wi-Fi infrastructure, present in most modern buildings. Our approach does not require any building modifications, network changes, or special equipment, such as sensors, to be deployed.
- **Facilitate a ready-to-use practical solution.** Our solution can be fully deployed by solely setting up a server that runs our software.
- **Allow for privacy-friendly detection.** Users' devices have to be identified for the purposes of people counting within certain building areas. However, our system does not reveal any information, other than the total number of persons and devices per building area, and what can be inferred from those.
- **Allow for a solution that comes at virtually zero cost.** Installation and management of our solution is simple, minimizing the costs of manpower necessary. The cost of setting up a server to run our application is considered negligible. In addition, our solution uses minimal network resources.

- **Demonstrate the accuracy of our solution with real-world experiments.** We implemented and deployed our approach in one of the main buildings of our university. In this setting, the majority of the building occupants are students and employees, who have a network account with which they log in the wireless network. We argue that this is a common situation in commercial and industry buildings, where the main occupants are either employees, or guests. Guests are often provided with access to the network via guests' accounts, which can be detected by our system in exactly the same manner as normal user accounts. To establish the ground truth, we physically monitored the areas where our system is deployed. Then, we evaluated our system results against the ground truth. Validation of our occupancy estimator results indicated an average accuracy of 88.74% (i.e., 88.74% of the persons actually being present in the monitored areas were also detected by our occupancy estimator).

2. Related Work

During the last decade, occupancy levels estimation within buildings, or rooms thereof has been extensively researched. Passive infra-red (PIR) networks is one of the approaches used to assist in occupancy estimation. In their seminal work, Dodier et al. (2006) were the first to deploy a network of infra-red sensors, and then apply Bayesian probability theory to estimate the occupancy within a building. Based also on data collected by passive infra-red sensors, together with the commonly used humidity, and CO_2 levels sensors, Han, Gao, & Fan (2012) showed that an Autoregressive Hidden Markov Model (ARHMM) performed better than the classical Hidden Markov Model (HMM), and Support Vector Machines (SVM), achieving an average estimation accuracy of 80.78%. Limitations of PIR networks include failure of detection of static humans, and sensitivity to external conditions. Additionally, in contrast to our work, PIR networks are specifically designed to determine occupancy.

Sensing environmental conditions to estimate occupancy has been also popular. Dong et al. (2010), deployed a complex wired and wireless sensor network measuring environmental conditions, and monitored the accuracy achieved by this setting with a wired camera network. The measured occupancy levels were properly adjusted, using Hidden Markov Models, to reflect reality more accurately. Evaluation of this work showed that they achieved a 73% accuracy on the number of people detected. Very similar results to the aforementioned have been presented earlier by Lam et al. (2009). In a somewhat different vein, although still measuring environmental conditions, Pan et al. (2014) proposed a system named BOES, for room-level occupancy estimation. Their approach is based on sparsely distributed vibration sensors, and achieves 85% accuracy in occupancy counting. The aforementioned techniques are intrusive, deploying complex, dedicated networks.

More works attempted to estimate occupancy based on sensing environmental conditions. Unlike the previously mentioned ones (Dong et al., 2010; Pan et al., 2014), these works (Ebadat et al., 2013; Yang et al., 2012) composed their experimental design in settings where existing sensors could be used, instead of deploying additional equipment. Ebadat et al. (2013) focused on occupancy estimation within rooms, rather than buildings, and exploited the information already collected by HVAC systems. They achieved an accuracy of up to 88.2%. Using also non-intrusive sensors for detection, and neural networks for data processing, Yang et al. (2012) reached up to 88% detection accuracy. We consider existence of environmental conditions sensors to be a strong assumption. While the majority of commercial, industrial, and public buildings have a basic Wi-Fi infrastructure, this is not the case for environmental monitoring sensors.

Occupancy detection based on video has also been proposed in the literature. Most works deploying video as a sensing method have been using additional methods to enhance the accuracy of their detection results. A combination of several sensing methods, including video, together with historical data on building utilization (Meyn et al., 2009), showed that combining approaches can significantly reduce the error range introduced by each approach separately. Meyn et al. (2009) managed to reduce the estimation error from 70%, in case of only HVAC measurements, to 11%. Tomastik et al. (2010) emphasized the importance of real-time data, as they focused on the emergency egress use-case. Their approach combined video together with an extended Kalman filter. Benezeth et al. (2011) developed an approach, based solely on video analysis of recordings by static cameras, which accomplishes human presence detection and activity characterization. However, we assess the infrastructure and management costs of a video-based sensing approach as high. In addition, video-based approaches are violating the occupants' privacy.

Innovative RFID-based occupancy detection methodologies are an indispensable part of the literature, as well. However, they are highly intrusive, requiring occupants to carry RFID tags. A novel approach based on RFID technology was proposed by Li, Calis, & Becerik-Gerber (2012). Focusing again on the use-case of demand-driven HVAC applications, they also aimed their attention to moving occupants. While they achieved a detection accuracy of 88% and 62% for stationary and mobile occupants, respectively, the proposed technology requires tracking tags to be attached to the occupants. To address positioning and tracking within construction environments, Woo et al. (2011) also used RFID technology.

In an attempt to avoid being intrusive, and remove the need for dedicated sensing equipment, Liao, & Barooah (2010) focused on occupancy modelling. They set up an agent-based model to simulate the behaviour of the occupants in commercial buildings. However, in the same manner as sensing alone could not produce highly accurate results, modelling alone could not either. Liao, & Barooah (2010) needed survey or sensor data on the building usage, for their model to produce reliable results.

Closer to our approach, Wi-Fi infrastructure has been used to address the occupancy estimation challenge. Existing solutions are, nonetheless, quite intrusive. Using smartphone capabilities and build-in sensors, Khan, Hossain, & Roy (2015) recently proposed an occupancy estimation and localization system, based on acoustic, accelerometer, and location sensors of the smartphone, though requiring active user collaboration for setting up this system. Balaji et al. (2013) use the Wi-Fi infrastructure within buildings, and collect data from the users' smartphones to determine occupancy, and adjust HVAC accordingly. However, they assume regular building occupants, whose privacy is not of concern. In contrast to our work, Balaji et al. (2013) also exclude building areas that are shared spaces, assuming they are constantly occupied, and they only answer the occupied/not-occupied question, without attempting to estimate the count occupancy. A combination of Wi-Fi together with other measurements (e.g., power consumption) is used by Christensen et al. (2014) to determine not only occupancy, but also number, identity, and activity of the persons in a building area. An interesting new approach (Depatla, Muralidharan, & Mostofi, 2015) attempts to estimate the number of people in a certain area, by placing two stationary antennas and measure the effect of crowds on the Wi-Fi signal, based on power. Unlike our solution, the latter two works are intrusive; both in terms of occupant's privacy (Christensen et al., 2014), and in terms of building modifications and supplementary equipment required (Depatla, Muralidharan, & Mostofi, 2015) to estimate occupancy.

Paper	Count Occupancy	No Dedicated Equipment	No Occupant Cooperation	Accuracy
Dodier et al., 2006	**X**	**X**	✓	-
Meyn et al., 2009	**X**	**X**	✓	79-89%
Benezeth et al., 2011 (a)	**X**	**X**	✓	97%
Woo et al., 2011	**X**	**X**	**X**	-
Balaji et al., 2013	**X**	✓	✓	86%
Pan et al., 2014 (a)	**X**	**X**	✓	99.55%
Tomastik et al., 2010	✓	**X**	✓	-
Dong et al., 2010	✓	**X**	✓	73%
Liao, & Barooah, 2010	✓	**X**	✓	-
Benezeth et al., 2011 (b)	✓	**X**	✓	83-93%
Yang et al., 2012	✓	HVAC	✓	65-88%
Han, Gao, & Fan, 2012	✓	**X**	✓	80.78%
Li, Calis, & Becerik-Gerber, 2012	✓	**X**	**X**	62-88%
Ebadat et al., 2013	✓	HVAC	✓	88%
Pan et al., 2014 (b)	✓	**X**	✓	85%
Christensen et al., 2014	✓	✓	✓	-
Depatla, Muralidharan, & Mostofi, 2015	✓	**X**	✓	-
Khan, Hossain, & Roy, 2015	✓	✓	**X**	80%
Our work	✓	✓	✓	88.74%

Table 1. Related work features' comparison.

In Table 1, we summarize the main features of the reviewed related work, and compare them to our work. Table 1 divides the related work into two main parts, separating the works answering solely the occupancy question (i.e., is the space in question occupied or not), from those estimating count occupancy (i.e., number of people in the space in question). As indicated in the third column of Table 1, unlike ours, most works require dedicated

equipment, or modification of existing ones to function. Some works (Ebadat et al., 2013; Yang et al., 2012) consider HVAC sensors as already existing infrastructure, which we regard as a strong assumption. Requiring occupants' cooperation is not common in such systems, as seen in the fourth column, but we have considered it an important feature. This is because any kind of occupants' cooperation is a strong requirement. In the last column of Table 1 we refer to the achieved accuracy of each proposed system, when available.

3. Wi-Fi-Based Occupancy Estimator

System Design

For the data collection our system relies on the Simple Network Management Protocol (SNMP). SNMP is a widely used protocol designed to manage and monitor network resources. Its main components are the managed devices (e.g., routers); the software running on these devices, called agent; and the software running on the manager (e.g., WI-Fi controller), called Network Management Station (NMS). The useful information we collect comes from the so-called SNMP authentication traps, which are sent from the Wi-Fi controller to our server. Specifically, we collect IP addresses, MAC addresses, and Usernames. To achieve an accurate count occupancy estimation for each building part, we have developed several SNMP trap filters.

Prerequisites. Our system is able to identify unique devices, and unique users carrying these devices, only for users who get connected to the specific Wi-Fi network under investigation. We conducted the validation experiments for our system in the premises of Saxion University of Applied Sciences. Three types of Wi-Fi services are offered by the university, namely the Eduroam network, the Saxion network, and the Guest network. All three types work identically, as far as SNMP authentication is concerned, and allow us, therefore, to estimate the number of users in the areas under investigation. Users carrying no device, or carrying devices that do not connect to any of the aforementioned three sub-networks, cannot be identified by our application.

To estimate the location of the identified users we assume a map of the complex, where our estimator should function, indicating the location where each Access Point is installed, together with the MAC address of the Access Point in question. For our experimental setup we have received an incomplete map of the university premises, which we completed upon visual inspection combined with application testing.

User Privacy. For the purposes of identifying unique users within the building under investigation, determinism is essential. Nevertheless, user privacy is of high importance. To satisfy both requirements while maintaining efficiency, we chose for hashing the collected usernames, and MAC addresses. For hashing we used the current state-of-the-art: SHA256. IP addresses are collected in plaintext for usability reasons. We do not consider this an infringement of privacy, because a user receives a new, random IP address after each authentication.

Although user information is anonymized before getting registered to our database, hashes are deterministic. This can potentially allow a malicious user to analyse that information over time, and determine user identities. To minimize the possibility of such an event, all data

entries are removed when the device is considered to have left the building, or after one hour of inactivity.

Multi-device User Filtering. It is possible for a user to carry more than one Wi-Fi enabled devices that connect to the network. We can filter this information, and therefore count separately number of unique users versus number of unique devices, based on the hashed username. To accurately locate a user within a building, smartphones are preferred over laptops. We assume that a user is more likely to carry a smartphone, and leave behind a laptop, when moving away from a working space. In addition, our exploratory experiments with different types of devices and operating systems demonstrated that Windows laptops by default block incoming ping requests. Ping requests (i.e., querying a computer on a network to determine whether there is a connection to it) are necessary for the other filtering techniques we have developed.

Thus, to determine which device of a user will serve for his/her unique identification within the building, we take two steps. First, we send a ping request to all devices of the same user. If only one device responds to the ping requests, our first filter associates the user with that device. If more than one device responds to the ping requests, a second filter checks which device's timestamp is more recently updated (because of a re-authentication request), and associates the user with that device. We chose for this functionality because our pilot experiments demonstrated that laptops send re-authentication requests approximately every 45-120 minutes, while smartphones approximately every 3-5 minutes. If no device responds to the ping requests, the one with the most recent timestamp will be associated with the user.

Database Management and Device Monitoring

Whenever a device gets authorized on an Access Point a database row concerning this device will be added or updated. The information in this row is: the IP address, and the hashed MAC address of the device, the MAC address of the Access Point, a timestamp, and the hashed username. We distinguish amongst the following events, initiated by our application, or the wireless device:

New Device Gets Authorized. If a device that is not currently registered in our database gets authorized, we insert a row in our active devices database, following the filtering procedures described in the paragraph *Multi-device User Filtering*.

Existing Device Gets Re-authorized. If a device already registered in our active devices database gets re-authorized, we update the MAC address of the Access Point, indicating the location of the device within the building, and the timestamp of the corresponding row of the database.

Reachable Devices Status Monitoring. We define as reachable devices the ones that do not block ping requests. At pre-set time intervals (currently set to two minutes) our application initiates the status monitoring process for these devices. First, the timestamp of each reachable device is checked. The timestamp indicates for how long the device has been observed to be inactive. If the timestamp is older than a pre-set time interval (currently set to ten minutes), a ping request is being sent to the device. If the device responds to the ping request, the corresponding timestamp is being updated. If the device no longer responds to

the ping request, all database entries regarding this device are being deleted, and the relevant counters are being updated.

Unreachable Devices Status Monitoring. We define as unreachable devices the ones that block ping requests. These are Windows laptops with the default firewall settings, and any other device that has been actively configured to block ping requests. At pre-set time intervals (currently set to two minutes) our application checks the timestamp of each unreachable device. If the timestamp for such a device is older than a pre-set time interval (currently set to 45 minutes), which indicates the minutes of observed inactivity, all information about this device is removed from the database, and the corresponding counters are being updated.

Blacklisting. After experimentation, we noticed that certain devices connected to the network never move from Access Point to Access Point, while they remain connected for long periods, often more than two working days in a row. These devices are blacklisted by a filter we have developed. A device enters the blacklist after being active, but not moved for a pre-set time interval. The device gets removed from the blacklist once it moves to another Access Point, or disconnects from the network. Blacklisted devices cannot be associated to users, and therefore do not contribute to the counting of unique users in the building.

Configurability

It is possible to configure our occupancy estimator, based on the requirements and resources available at the setting where it is to be deployed. One parameter that can be configured is the ***Wi-Fi event retention time*** in the database. To prevent the creation of user profiles by malicious parties that can eventually reveal user identities, we have set this value to 60 minutes. Thus, any database entry with a timestamp older than 60 minutes, is being deleted.

It is also possible to configure the ***unreachable users retention time***. This value can be set to 0, if the desired retention time for unreachable users is identical to the Wi-Fi event retention time. Otherwise, we can adjust this value by selecting a value lower than the one for the Wi-Fi event retention time. Our initial pilot experiments demonstrated that 45 minutes is a suitable value for this parameter, as unreachable devices (mainly Windows laptops) often send one re-authentication request within this time-frame. Thus, we have set the default value of this parameter to 45 minutes.

To keep the accuracy of our occupancy estimator high, we monitor in pre-set time intervals whether the reachable users are still within the Wi-Fi range, and therefore within the building area. We do so by sending them ping requests. This is necessary, as leaving devices do not send any Wi-Fi events that we can monitor, and update our counts accordingly. It is possible to configure ***the reachable users probe time interval***, and the value to be set depends on the size of the network under investigation. For our experiments we have set this value to 10 minutes. We recommend a value $\geq$ 10 minutes for networks that serve more than 5000 users. This is to avoid overloading the network with ping requests.

Based on the maximum time that a typical occupant spends in the area where our occupancy estimator operates, we can configure the ***blacklisting time interval***. Recall that blacklisted devices are the ones that remain connected to the Wi-Fi network, but stationary for a pre-set time interval. The main university building is open from Monday to Thursday 7.30-22.00,

and on Fridays 7.30-17.30. We assume that no occupant stays in the building for the full 14.5 hours that the building is available, at most. Hence, we selected the default value for this parameter to be 14 hours.

For each of the aforementioned three parameters (unreachable users time interval, reachable users probe time interval, and blacklisting time interval) there is an additional parameter: the ***check initiation time interval***. This parameter defines how often the application will initiate each of the three condition checks. For unreachable users, and reachable users probe time intervals we have set as default values two minutes for both. This value has to remain low to keep the occupancy estimator as up-to-date as possible. However, our pilot experiments showed that 1 minute value or lower can cause performance issues to our application. For the blacklisted devices we set the default value to 15 minutes, to keep it proportional to the blacklisting time interval, and for performance reasons.

Data Reporting

Our live monitoring system, collects the data as discussed in the *System Design* Section, and updates our database. This allows us to report near real-time on certain values, including the estimated number of unique users in the building, only by querying our database. An example of the overall data reporting functionality of our application, while our system is running in our testing environment, is depicted in Figure 1. The values our application reports on are as follows:

- **Number of unique users:** The number of unique users is equal to the number of unique (hashed) usernames in our database.
- **Number of unique devices:** The number of unique devices on the network is identical to the total number of entries in our database.
- **Number and percentage of reachable devices:** This is the number and percentage of devices in our database that respond to incoming ping requests.
- **Number and percentage of unreachable devices:** This is the number and percentage of devices in our database that block incoming ping requests.
- **Percentage of users having been active within a certain time interval:** This is the percentage of all users (reachable and unreachable) that has been observed as active in the latest pre-set number of minutes. Active users are the ones having changed Access Point within the pre-set time interval, or having responded to a ping request, or having initiated an SNMP re-authentication. For our experiments we have set the time interval to 15 minutes.
- **Percentage of unreachable users having been active within a certain time interval:** This is the percentage of the unreachable users, by means of sending ping requests, who nevertheless have been observed as active (e.g., by having changed Access Point) within the latest pre-set number of minutes. This value was also set to 15 minutes during our experiments.
- **Number of devices per user:** This indicates the average number of unique, Wi-Fi enabled, and active devices for each user that carries at least one such device.
- **Number of blacklisted devices:** This is the number of devices having been observed as active, but remain stationary for more than 14 hours in a row.
- **Number of Access Points in use:** The number of Access Points in use, gives an indication of the parts of the building that are occupied by users.

In addition, our application reports on the number of unique users in a more granular manner:

- **Number of users per building part:** We have also developed an interface, which allows the application manager to select a specific building of the whole complex, then a floor thereof, and then the exact room of interest, and have the number of users in that room displayed. This also displays the number of unreachable users out of the total number of connected ones to the specific Access Point.

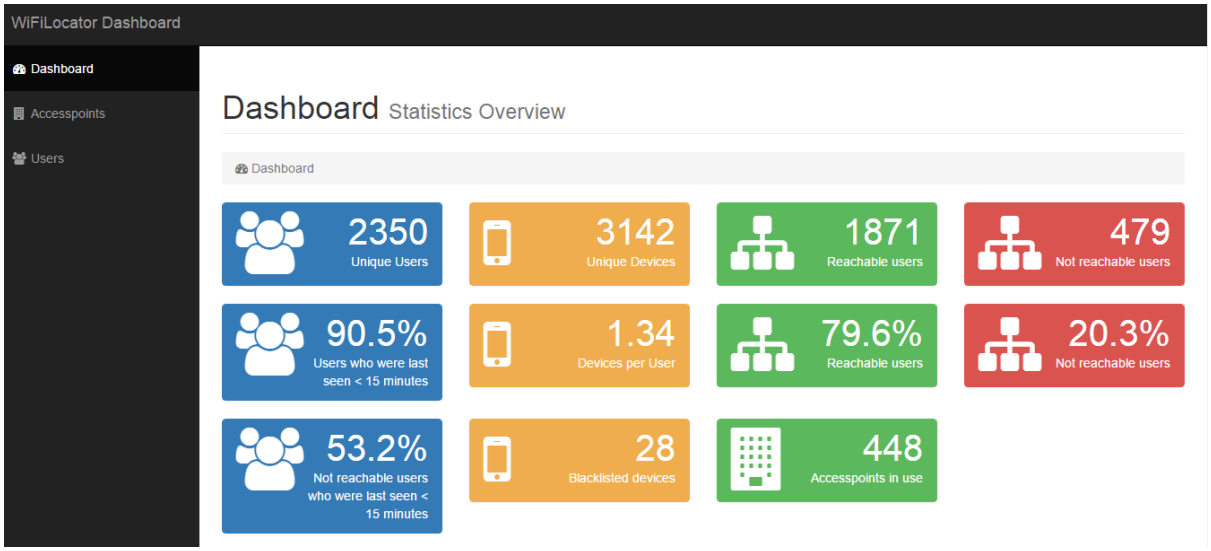

Figure 1. Overall Data Reporting.

In Figure 2 we present a screenshot of our application, where one of the five main buildings of the complex we conducted validation, named Wolvecamp, has been selected. Then Floor 1 out of the six floors of this building was selected by the application operator. After submitting the search, the application reports on the total number of users on the selected floor. The application also reports on the number of users per room (identified by the Access Point installed in the room in question), and the number of connected users in this room that are unreachable (i.e., users probably having a stationary Windows device).

4. Validation

Experimental Design

We have validated our results by visually inspecting selected building areas, where the application is concurrently performing people counting estimation, and then comparing the outcomes. More precisely, at the beginning of each experimental session the total number of people in the area under inspection is being counted. Each experimental session (time slot) lasts between one and two hours. Researchers are monitoring the entrances/exits of the area under investigation, and report on the number of incoming and outgoing people. Every five minutes, the researchers report on their results, and take note of the results concurrently reported by the occupancy estimator. This is how we establish the ground truth, and compare it with the outcomes of our application.

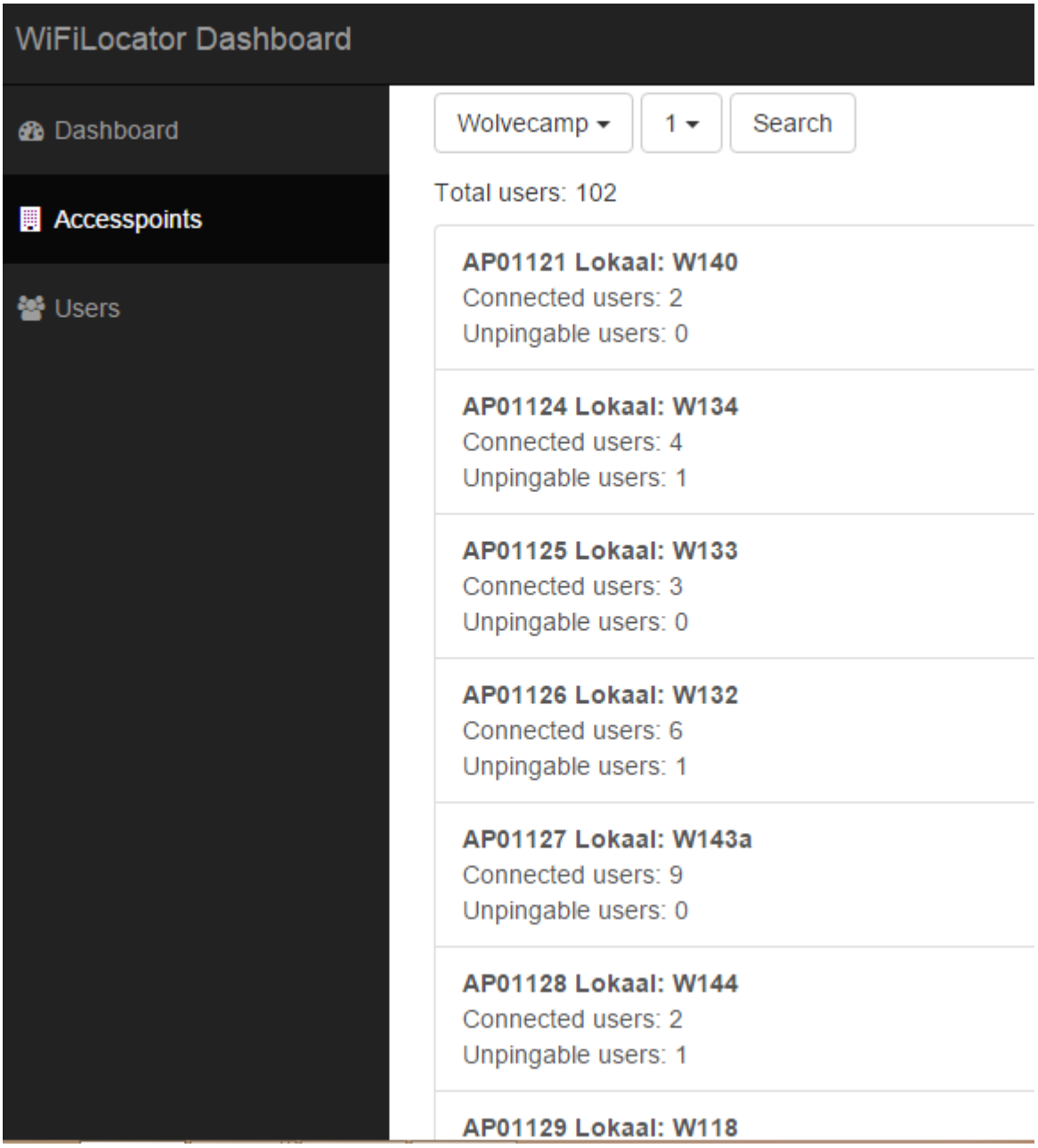

Figure 2. Granular Data Reporting.

We focused on testing during the busiest hours, in terms of number of people being present at the university. Our earliest experiments started at 9.40, while the latest ones finished at 15.50. To enhance the validity of our results, we conducted the experiments in different time slots, within the aforementioned selected times. For cross validation purposes, we have conducted the experiments in two different locations within the university complex. An overview of one of the experimental locations is shown in Figure 3.

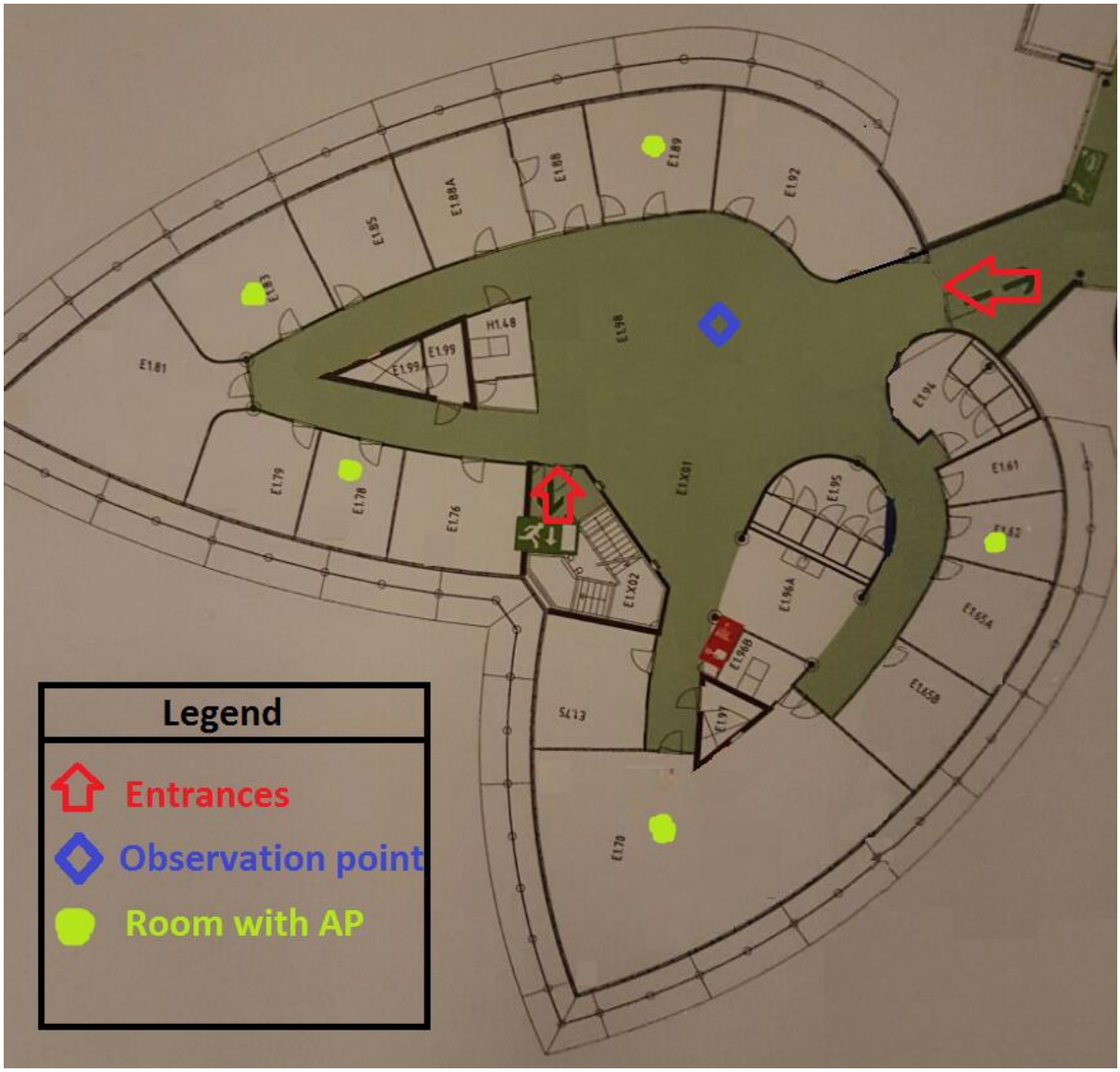

Figure 3. Testing Location: Eldering building; Floor 1; Right wing.

Experiment Results

We have collected the results of two sets of experiments, conducted in two different locations within the main our university complex. The first set of experiments was conducted on the third floor of the Wolvecamp building in the main complex of Saxion Enschede. The experiments took place between the 9th and 10th of December 2015, and between the 14th and 15th of December 2015. We have collected 98 observations from that location in total by visually inspecting the area under investigation, and comparing our findings with the application results every 5 minutes. During these experiments we have visually counted 35 individuals at minimum, and 263 individuals at maximum (ground truth). At the same time, our application has counted 35 individuals at minimum, and 241 individuals at maximum. The average number of people in the area under investigation during the whole experimentation period was 114.8 individuals, as counted by the visual inspection, and 101.5 individuals, as counted by our occupancy estimator.

The greatest change in the number of people observed within 5 minutes was 59 individuals (81 incoming individuals, and 22 outgoing). Our application managed to identify

immediately (i.e., at the 5 minute interval of comparison) 39 individuals as having left, and eventually reached an accuracy of 94.1% again within 10 minutes after this change. Note that within these extra 10 minutes another 98 individuals entered, and 59 individuals left the area of experimentation. The 98 observations for the specific location have been summarized in Figure 4.

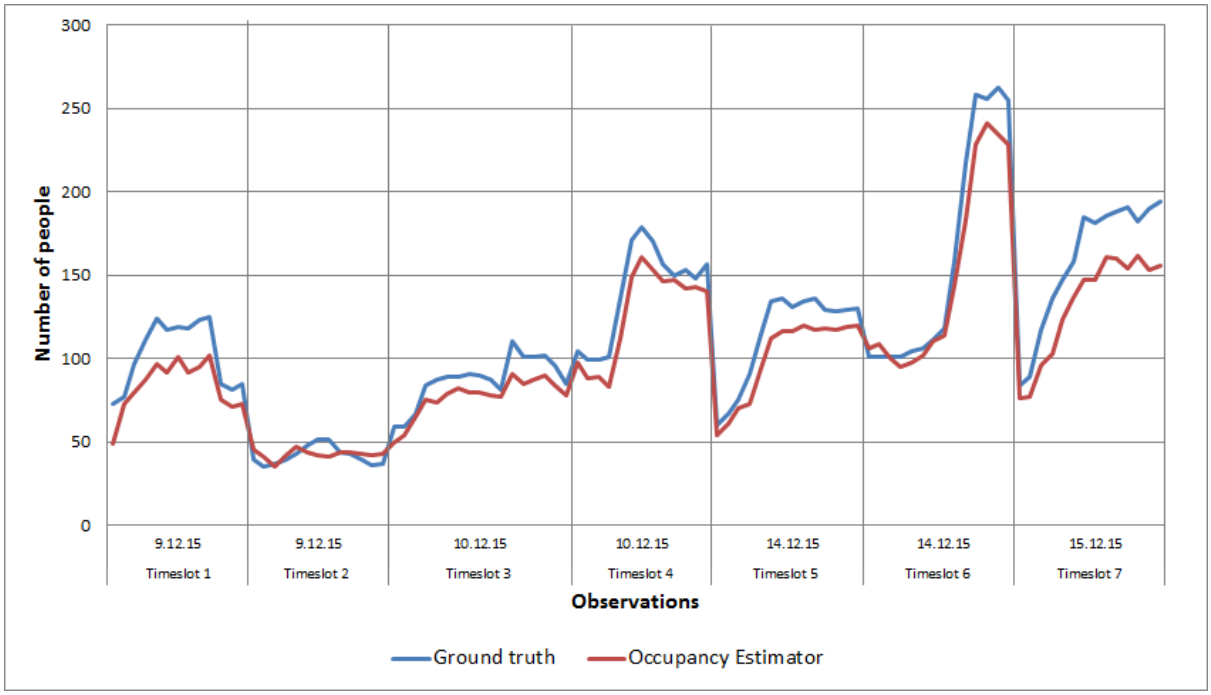

Figure 4. Experiment Results Wolvecamp building.

The accuracy of our occupancy estimator varies from 67.12% in the worst observation, and it reaches up to 100% accuracy. On average, the occupancy estimator achieved 87.86% accuracy at this location. The accuracy of each of the 98 observations in this building area, sorted by accuracy, has been plotted in the chart of Figure 5. In 90 out of the 98 observations (91.8%) the occupancy estimator demonstrated accuracy above 80%. In 36 out of the 98 observations (36.7%) the accuracy of the estimator reached and exceeded 90%.

The second set of experiments was conducted at the right wing, on the first floor of the Eldering building in the main complex of Saxion Enschede. The experiments were executed between the 16th and 18th of December 2015, between the 11th and 15th of January 2016, and on the 18th and 20th of January 2016. During this period we have collected 199 observations. The minimum number of individuals counted visually was 28, and the maximum 88 individuals. At the same time, our occupancy estimator has counted 27 individuals at minimum, and 81 individuals at maximum. The average number of people in the area under investigation during the whole experimentation period was 56 individuals, as counted by the visual inspection, and 53.5 individuals, as counted by our application. The 199 observations concerning this set of experiments have been summarized in Figure 6.

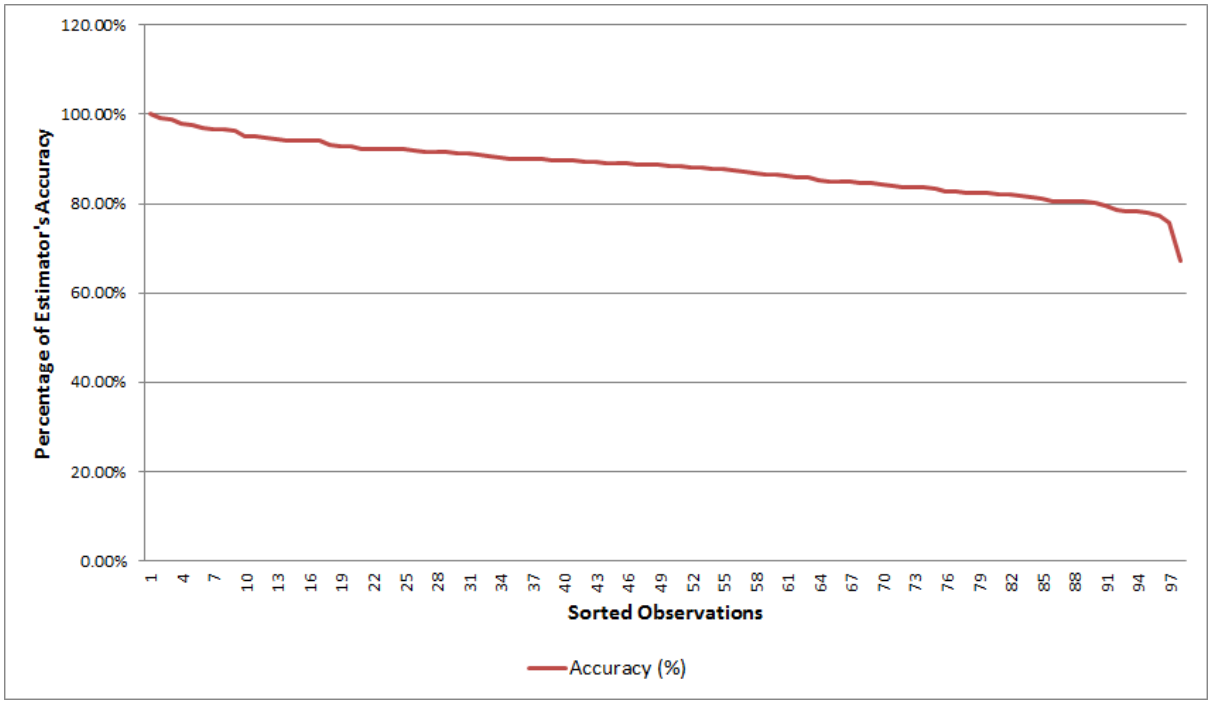

Figure 5. Estimator's Accuracy Wolvecamp building.

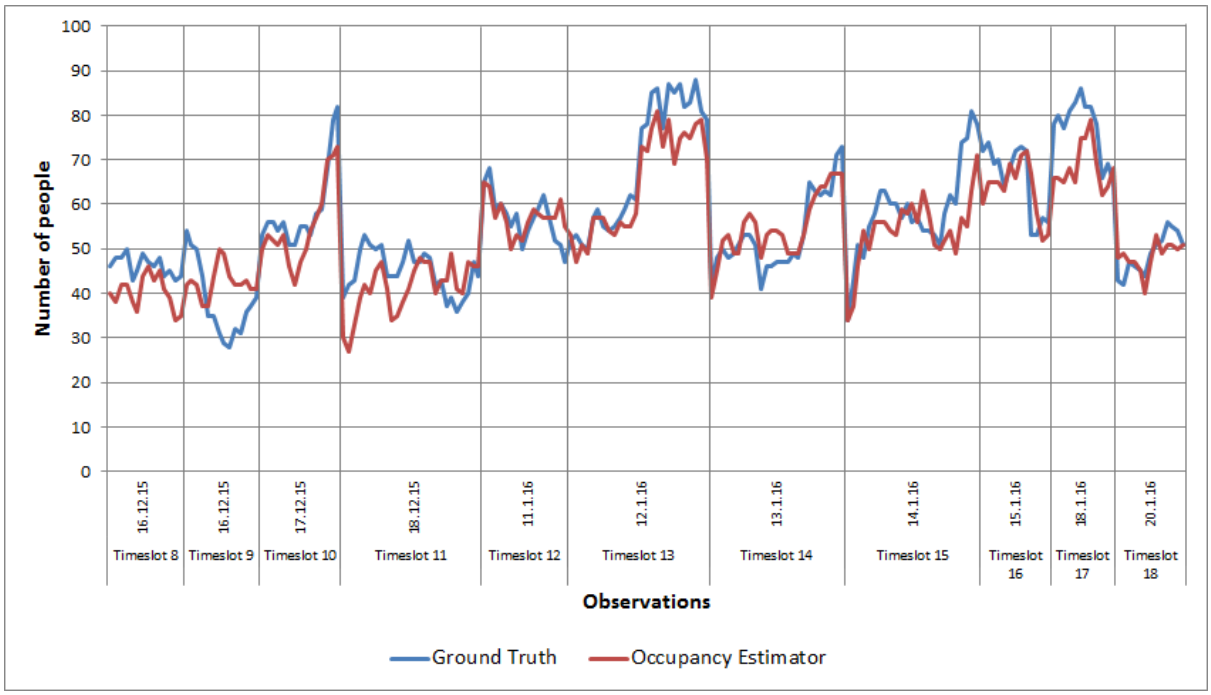

Figure 6. Experiment Results Eldering building.

The accuracy of our occupancy estimator varies from 31.03% in the worst observation, and it reaches up to 100% accuracy in 15 observations. On average, the occupancy estimator achieved 89.62% accuracy in the Eldering building. We have sorted the 199 observations based on the accuracy demonstrated by the occupancy estimator, and plotted them in the chart displayed in Figure 7. In 174 out of the 199 observations (87.4%) the accuracy of the occupancy estimator was 80% or higher. 60.8% of the observations (121 observations) indicated a 90% or higher accuracy of the occupancy estimator, with 7.5% of the observations being 100% accurate.

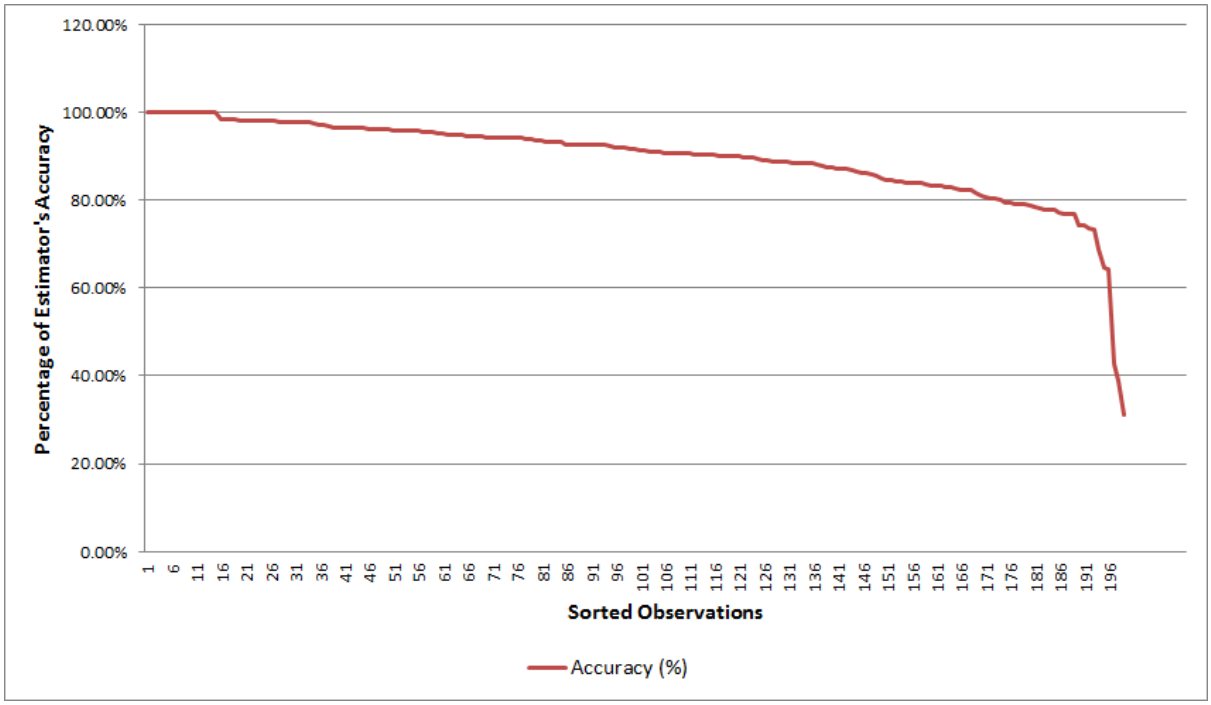

Figure 7. Estimator's Accuracy Eldering building.

In an attempt to make optimal use of the data gathered by our occupancy estimator, we have also reported on the number of unreachable users during the last set of experiments. The additional information has been reported for the period between the 12th and the 15th of January 2016, and on the 18th and 20th of January 2016. This accounts for 114 observations in total. The results for this period, including the number of unreachable users, have been mapped in Figure 8. The number of unreachable users (i.e., the fraction of the occupancy estimator's results that do not respond to ping requests) grows proportionally to the total number of occupants, and it does not demonstrate patterns similar to the error measured by the occupancy estimator.

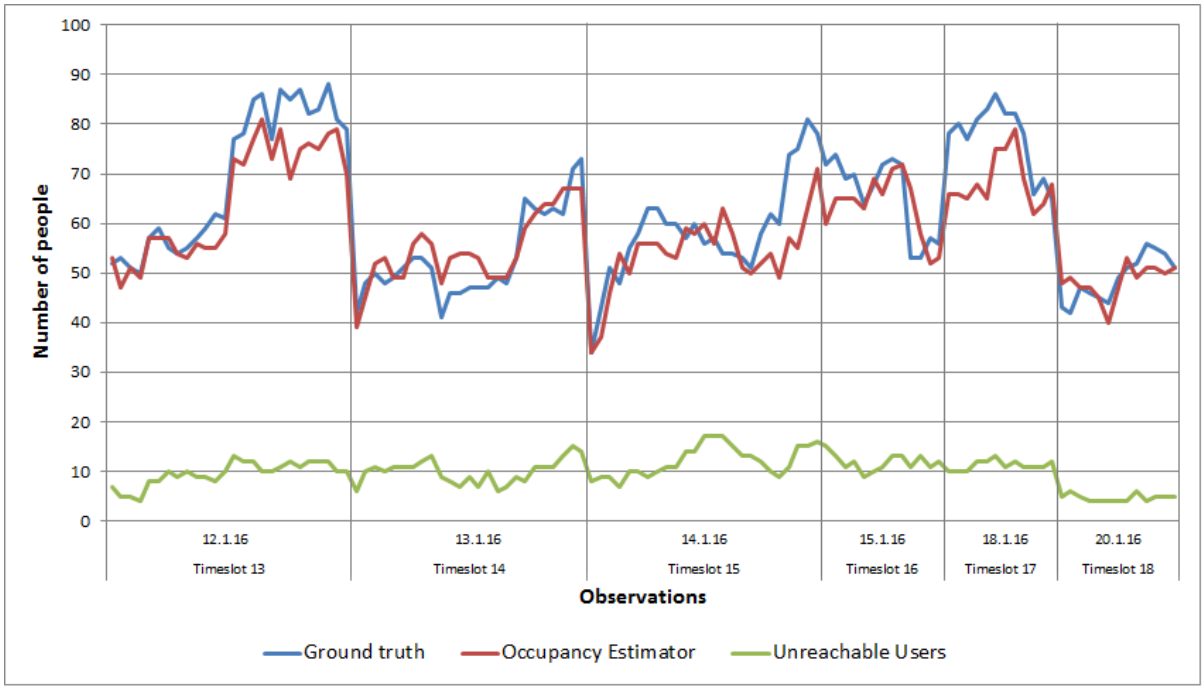

Figure 8. Experiment Results including unreachable users.

5. Discussion

Limitations

Clearly, our occupancy estimator can only detect users carrying a Wi-Fi enabled device that is connected to the Wi-Fi network of the area, where occupancy count is of interest. Recall also that unlike the visual inspection results, the occupancy estimator's function is near real-time, not truly real-time, which results in timing differences.

One of the greatest challenges for our occupancy estimator was to detect leave events. A device that is granted access to the network is instantly detected upon authentication. A device that moves within the building complex premises gets almost instantly detected, moving from one Access Point to another. On the other hand, a device that leaves the area under investigation sends no network events, making detection of leaving events more demanding. Therefore, it is not uncommon for the occupancy estimator to output a count result that is higher than the actual number of people in the area under inspection. Note, however, that the occupancy estimator's counts are, in the majority of the observations, equal or lower than the visual inspection's counts (establishing the ground truth). Overall, our estimator can be measuring error, especially for unreachable devices, which can be out of range, but appear in our estimator results as still connected for a maximum of 45 minutes.

Establishing the ground truth was non-trivial. The high number of incoming and outgoing individuals made visual inspection and reporting challenging, even with four researchers counting simultaneously at the same location. Individuals may have been unidentified in the establishment of ground truth, for being not visible to the researchers performing the counting (e.g., people being in the toilets when the initial total counting for the area was taking place). When the researchers were uncertain about their own counting of incoming/outgoing individuals, the total number of people in the area was counted anew, and the results were adjusted accordingly in the new calculations.

Conclusion

We have developed a plug and play solution to address the problem of occupancy estimation. Our proposal, being solely software based, requires no building modifications, or additional equipment to be deployed. Assuming basic Wi-Fi network infrastructure, present in any modern building nowadays, we can set up a server that runs our software, and have an accurate occupancy count estimation. Our solution does not drain the network resources, and we therefore argue that it comes at virtually zero cost. In addition, it is absolutely non-intrusive, not only as far as building modifications are concerned, but also regarding occupants privacy. Validation of our occupancy estimator's results indicated 88.74% average accuracy.

Our system alone can be used to determine count occupancy so as to allow HVAC applications to automatically adjust the indoors environmental conditions, and send near real-time occupancy information to local emergency services. The proposed occupancy estimator can also be used as an additional input to multiple input modelling systems, used to predict the number of occupants in a specific building part. In such a case, the results produced by

our occupancy estimator can be highly weighted input variables, given the accuracy that our tool demonstrates.

Future Work

An interesting future direction to improve our system itself, is to study other relevant variables, and how these can be used in a mathematical model to achieve an even higher occupancy estimation accuracy. These variables can be related to the number of currently active users, or the fraction of currently active users out of the ones that are reported as unreachable.

Another interesting direction is to experiment with our occupancy estimator both in different types of environments (e.g., commercial buildings, and companies), as well as experiment on a larger scale. Although we consider our validation setup to be realistic, significant performance differences might come into sight if validation is to take place in a different setting, or with a considerably larger occupants population. Slight modifications on the Wi-Fi network infrastructure of a building, would also allow us to get access to a greater number of variables, which would likely increase the accuracy of the proposed solution. Our goal was to address occupancy estimation at zero cost, using only existing infrastructure, and by proposing a ready-to-use solution. Solutions very similar to the proposed one, although somewhat more intrusive, can be deployed for a more accurate occupancy count estimation.

Finally, we consider as an interesting future path, the use of our proposed solution as a highly weighted input variable of a modelling system, combining various environmental measurements. Validation results of such a combination are of high importance, as they are likely to enhance the accuracy of the combined proposed solutions.

Acknowledgment

This work has been done in the context of TEC4SE project, which is supported by Thales NL and Veiligheidsregio Twente. The authors would like to thank Elmer Lastdrager for his feedback on earlier versions of this manuscript. We would also like to thank the lectorate Risicobeheersing of Saxion, and especially lector Wilbert Rodenhuis, and Jeroen Neuvel.

References

Balaji, B., Xu, J., Nwokafor, A., Gupta, R. , & Agarwal, Y. (2013). Sentinel: Occupancy Based HVAC Actuation Using Existing Wi-Fi Infrastructure within Commercial Buildings. In Proceedings of the 11th ACM Conference on *Embedded Networked Sensor Systems, page 17*.

Benezeth, Y., Laurent, H., Emile, B., & Rosenberger, C. (2011). Towards a Sensor for Detecting Human Presence and Characterizing Activity. *Energy and Buildings, 43*(2), 305-314.

Christensen, K., Melfi, R., Nordman, B., Rosenblum, B, & Viera, R. (2014). Using Existing Network Infrastructure to Estimate Building Occupancy and Control Plugged-in Devices in User Workspaces. *Communication Networks and Distributed Systems, 12*(1), 4-29.

Depatla, S., Muralidharan, A., & Mostofi, Y. (2015). Occupancy estimation using only Wi-Fi power measurements. *IEEE Journal on Selected Areas in Communications, 33*(7), 1381-1393.

Dodier, R. H., Henze, G. P., Tiller, D. K., & Guo, X. (2006). Building Occupancy Detection through Sensor Belief Networks. *Energy and buildings, 38*(9), 1033-1043.

Dong, B., Andrews, B., Lam, K. P., Höynck, M., Zhang, R., Chiou, Y. S., & Benitez, D. (2010). An Information Technology Enabled Sustainability Test-Bed (ITEST) for Occupancy Detection through an Environmental Sensing Network. *Energy and Buildings, 42*(7), 1038-1046.

Ebadat, A., Bottegal, G., Varagnolo, D., Wahlberg, B., & Johansson, K. H. (2013). Estimation of Building Occupancy Levels through Environmental Signals Deconvolution. In Proceedings of the 5th ACM Workshop on *Embedded Systems For Energy-Efficient Buildings, pages 1-8*.

Han, Z., Gao, R. X., & Fan, Z. (2012). Occupancy and Indoor Environment Quality Sensing for Smart Buildings. In Proceedings of the IEEE International Conference on *Instrumentation and Measurement Technology (I2MTC), pages 882-887*.

Khan, M. A. A. H. , Hossain, H. M., & Roy, N. (2015). Infrastructure-less Occupancy Detection and Semantic Localization in Smart Environments. In Proceedings of the 12th International Conference on *Mobile and Ubiquitous Systems: Computing, Networking and Services, pages 51-60*.

Lam, K. P., Höynck, M., Dong, B., Andrews, B., Chiou, Y. S., Zhang, R., Benitez, D., & Choi, J. (2009). Occupancy Detection through an Extensive Environmental Sensor Network in an Open-Plan Office Building. *IBPSA Building Simulation, 145*, 1452-1459.

Li, N., Calis, G., & Becerik-Gerber, B. (2012). Measuring and Monitoring Occupancy with an RFID Based System for Demand-Driven HVAC Operations. *Automation in construction, 24*, 89-99.

Liao, C., & Barooah, B. (2010). An Integrated Approach to Occupancy Modelling and Estimation in Commercial Buildings. In Proceedings of the IEEE *American Control Conference (ACC), pages 3130-3135*.

Meyn, S., Surana, A., Lin, Y., Oggianu, S. M., Narayanan, S., & Frewen, T. A. (2009). A Sensor-Utility-Network Method for Estimation of Occupancy in Buildings. In Proceedings of the 48th IEEE Conference on *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC , pages 1494-1500*.

Pan, S., Bonde, A., Jing, J., Zhang, L., Zhang, P., & Noh, H. Y. (2014). BOES: Building Occupancy Estimation System Using Sparse Ambient Vibration Monitoring. *SPIE Smart*

Structures and Materials+ Nondestructive Evaluation and Health Monitoring, pages 90611-90611.

Stottelaar, B., Senden, J., & Montoya, L. (2014). Online social sports networks as crime facilitators. *Crime science*, *3*(1), 8.

Tomastik, R., Narayanan, S., Banaszuk, A., & Meyn, S. (2010). Model-Based Real-Time Estimation of Building Occupancy During Emergency Egress. *Pedestrian and Evacuation Dynamics 2008*, pages 215–224.

Woo, S., Jeong, S., Mok, E., Xia, L., Choi, C., Pyeon, M., & Heo, J. (2011). Application of Wi-Fi-Based Indoor Positioning System for Labor Tracking at Construction Sites: A Case Study in Guangzhou MTR. *Automation in Construction*, *20*(1), 3-13.

Yang, Z., Li, N., Becerik-Gerber, B., & Orosz, M. (2012). A Multi-Sensor Based Occupancy Estimation Model for Supporting Demand Driven HVAC Operations. In Proceedings of the 2012 International Symposium of Society for Computer Simulation on *Simulation for Architecture and Urban Design, pages 49-56*.