Spatio-temporal Facility Utilization Analysis from Exhaustive WiFi Monitoring

Thor S. Prentow[a],
Antonio J. Ruiz-Ruiz[b], Henrik Blunck[a], Allan Stisen[a], Mikkel B. Kjærgaard[a]

[a]{prentow,blunck,stisen,mikkelbk}@cs.au.dk,
Department of Computer Science, Aarhus University, Aabogade 34, 8200 Aarhus N, Denmark
[b]*antonioruiz@um.es, Department of Computer Engineering,*
University of Murcia, Campus de Espinardo, Campus de Espinardo, CP 30100, Murcia Spain

---

Abstract

The optimization of logistics in large building complexes with many resources, such as hospitals, require realistic facility management and planning. Current planning practices rely foremost on manual observations or coarse unverified assumptions and therefore do not properly scale or provide realistic data to inform facility planning. In this paper, we propose analysis methods to extract knowledge from large sets of network collected WiFi traces to better inform facility management and planning in large building complexes. The analysis methods, which build on a rich set of temporal and spatial features, include methods for quantification of area densities, as well as flows between specified locations, buildings or departments, classified according to the feature set. Spatio-temporal visualization tools built on top of these methods enable planners to inspect and explore extracted information to inform facility-planning activities. To evaluate the proposed methods and visualization tools, we present facility utilization analysis results for a large hospital complex covering more than 10 hectares. The evaluation is based on WiFi traces collected in the hospital's WiFi infrastructure over two weeks observing around 18000 different devices recording more than a billion individual WiFi measurements. We highlight the tools' ability to deduce people's presences and movements and how they can provide respective insights into the test-bed hospital by investigating utilization patterns globally as well as selectively, e.g. for different user roles, daytimes, spatial granularities or focus areas.

Keywords: WiFi Monitoring, Facility Management, Spatio-temporal Data Analysis

1. Introduction

Healthcare administrators are constantly under pressure to reform the healthcare system organization by planning activities to better utilize available resources to minimize cost but at the same time offer a high quality healthcare service [1, 2]. The design and maintenance of a cost-effective and high quality healthcare system is an ongoing high-priority challenge for most governments around the world. A crucial part of this challenge is the difficulty inherent in planning hospital activities—as these require an accurate knowledge of the hospital environment, of the availability of resources (both materials and personnel), of knowledge about flows of personnel and patients, and usage of services and facilities. One example where better planning can help optimize healthcare services are removal of inefficiencies in patient flows, e.g., patient misplacement or late arrivals of patients, which result in surgery cancellations [3].

Today, only statistics from patient records are generally available to hospital facility planners [2], e.g. number of ambulant treatments and hospitalizations. Other existing approaches [4, 5] have tried to address the lack of knowledge using a modeling approach. These approaches focus on length of stay and flow of patients between departments to provide models reflecting the complex, variable, dynamic and multi-dimensional nature of hospital systems. However, in [6] the authors demonstrate that such model-based calculations typically do not provide the appropriate information needed to obtain reliable results—since the models do not take into account all variables influencing the continuous operations at a hospital. Examples of such variables include: i) amount and spatio-temporal distribution and flow of visitors—influencing the planning of offered facilities such as seating areas, parking spaces, and toilets; ii) precise up-to-date information about people within the building complex such as their role as patients, visitors, and staff.

Nowadays, widespread user devices such as smartphones, tablets and in the future also smart watches, emit WiFi signals on a frequent but irregular basis [7]. Moreover, the already available wireless infrastructures in large building complexes, like hospitals, enable the collection of large data sets of WiFi measurements that can be used not only to analyze the network's performance and usage, as proposed in earlier work among others [8, 9, 10], but potentially also the density and flow of people within the building. Compared to earlier approaches based on Bluetooth, in urban [11] or indoor settings [12], or based on video in indoor settings [13], the use of WiFi comes with lower setup costs, due to the existing deployment, for monitoring complete large-scale building complexes. However, analysis methods are missing that allow to extract information, relevant for planning, from collected large-scale WiFi data sets.

In this article, we extend our earlier work [14] proposing analysis methods to extract knowledge from large sets of WiFi traces to better inform facility planning in large building complexes. The analysis methods build on a rich set of temporal and spatial features extracted from the WiFi traces. The analysis methods include methods for i) noise removal, ii) quantification of people densities and flows at locations of interest and iii) analysis of traffic flow, both globally as well for individual foci, on e.g. specific user groups, departments, and/or daytimes. To remove noise we propose methods to clean data, filtering out, e.g., device traces that are close to the perimeter of the building complex but not within it. We do so by labeling these devices as beyond building-perimeter devices using machine learning-based classification with a novel set of features calculated from raw WiFi signal data. For estimating people densities and flows in areas we propose heuristics to filter streams of calculated device positions—assessing, among others, the number of enter and exit events. For traffic flow analysis between specific areas, we employ the defined feature sets as well as time-based filters to allow for a configurable flow analysis according to the needs of e.g. domain experts. The additions to the conference version of this article focus on the spatio-temporal visual tools for facility utilization analysis which are built on top of the described methods. Specifically, we present travel-based graphs as a basis to visualize traffic flow and how these allow to investigate and assess facility utilization, globally as well as selectively, e.g. for different user roles, daytimes, spatial granularities or focus areas.

To evaluate the proposed methods, we present results for a large hospital complex covering more than ten hectares in which we have collected WiFi traces over two weeks observing around 18000 different devices recording more than one billion individual WiFi measurements. Moreover, as background information we also present detailed statistics of the observed devices, e.g., type of devices and the frequency of observations. We present quantitative results for the analysis methods, e.g., for noise removal of beyond building perimeter devices where results demonstrate over 95% accuracy for correct removal. For the quantification of flows we present comparisons with manually recorded flows. Additionally, we present example visualizations such as heat and flow maps that both highlight the visualizations' potential as inspection tools for planners and provide interesting insights into the hospital's workings.

The presented methods can be generalized and thus applied not only to hospital settings but enable facility analysis also in other types of large building complexes such as industrial facilities, shopping malls or public buildings in general. The proposed methods can be also used to analyze the spatio-temporal distribution of people to offer better planing services and facilities, e.g., seating areas, parking spaces, toilets, and their maintenance, e.g., for cost-efficient scheduling of cleaning personnel at times of low load on the respective facilities.

2. Related work

Existing work utilizing measurements from wireless networks [8] focused on analyzing the networks' performance and usage. The analysis was based on aggregating the data into various forms of graphs and statistical summaries; for instance, to obtain statistics about the number of devices that made use of the network, which applications the network was used for, and the mobility of the users. The main aim of these studies was to improve the design, modeling and management of wireless networks in regards to, e.g., improved protocol designs or better adaptability for areas where APs exhibit a lot of network traffic. Such studies have been performed both in an university campus settings [8], corporate settings [9] and urban settings [10]. For a campus setting Calabrese et al. [15] proposed methods to explore overall user behavior for buildings on the campus but did not relate it to the within-building movements.

Another line of work has utilized data collected from people's own devices instead of using data from wireless networks. Such work has analyzed different aspects of people's behavior and of the places they visit. Chon et al. presented a system for

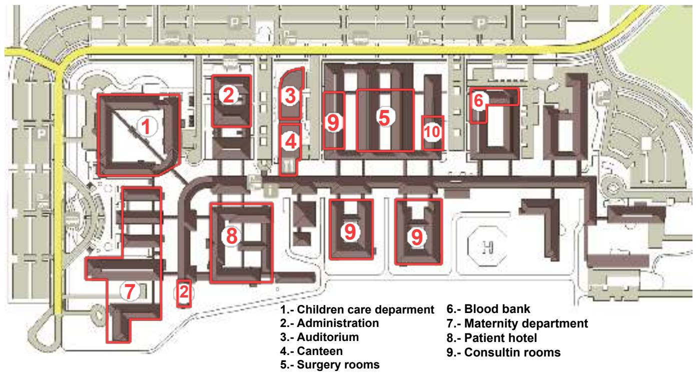


Figure 1: Aarhus University Hospital - Skejby complex.

categorizing places from mobile device data [16]. Vu et al. [17] presented a framework for constructing predictive models of people's movement. Focusing on sensing of the collective behavior of crowds, different methods have been proposed, e.g., to estimate properties regarding flocking, followers and density. Kjærgaard et al. [18, 19, 20] propose methods for flock detection and follower detection based on mobile sensing data. Neil et al. [11] consider methods for counting people in an urban setting using Bluetooth scanning. Other approaches focus on traffic analysis, including Musa et al. [7], and study vehicle tracking based on passive WiFi transmissions. The above study demonstrated that tracking unmodified devices using WiFi monitoring is feasible in outdoor settings but it did not consider indoor settings or facility planning. In contrast to previous work in this paper we propose analysis methods utilizing data from WiFi networks in large building complexes. These methods are designed to extract knowledge from such data to inform facility planning.

3. Hospital Testbed

During the process of developing the proposed analysis methods we have collaborated with staff from the planning and IT departments at Aarhus University Hospital. In discussions the staff told that their current practices for planning are mainly based on statistics from patient records and coarse estimates which is common according to existing research studies [2]. Furthermore, they were very interested in new means of obtaining and using more realistic information for their planning activities.

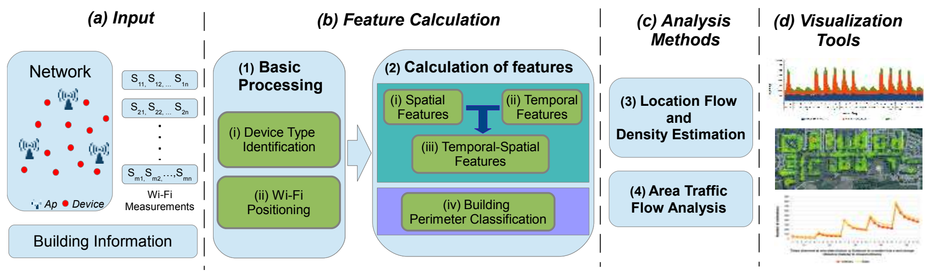


Figure 2: Overview of the steps involved in data processing feature extraction, analysis methods and visualization.

In collaboration with the hospital a data set of WiFi measurements was collected throughout the hospital complex. See Figure 1 for an overview of the complex. The hospital features 22 different buildings with up to 3 floors, covering an area of more than 10 hectares. The entire hospital relies on a wireless network infrastructure that covers all of its buildings, with the exception of those areas reserved to surgery rooms, where, due to safety reasons, electromagnetic radiation is restricted. The total amount of access points (APs) available in the hospital is 798, with most of them (around 95%) being Trapeze and Juniper devices. The network provides several virtual networks including a guest network open to the general public. The system architecture used for data collection is network-based, i.e., WiFi measurements are collected by the APs on all WiFi channels and forwarded to a central server which stores them to a database. Our data collection was carried out for 15 days using all available APs, collecting in total more than a billion of WiFi measurements from around 18000 different devices.

One important aspect in large-scale mining studies is that some of the extracted features (e.g. user position) are privacy sensitive—especially when working in hospital environments, since personal health information must be protected in regards to identification of individuals. Regarding this concern, we emphasize that we only collected network scan frames, and used an anonymization procedure during data acquisition that ensures a high level of privacy protection. Following the same approach as utilized for the Nokia data challenge [21], MAC addresses were encrypted by hashing after concatenating them with a secret key. This ensures that the collected tracking information can not be re-associated with a specific device.

4. Overview of Analysis Methods

Figure 2 illustrates how the proposed analysis methods build upon each other and together enable a tool chain to extract knowledge for facility planning and provide associated visualizations. The data used as input (a) are provided from two different sources: WiFi measurements from a large-scale wireless network, and a geometric model of the perimeter of the building complex. The feature calculation phase (b) is divided into two steps consisting of: (1) basic processing where the type of the device is identified (1.i) and the raw WiFi measurements are converted to positions using existing WiFi positioning algorithms (1.ii); and (2) calculation of a rich set of spatial (2.i), temporal (2.ii) but also spatio-temporal (2.iii) features to enable the analysis methods. The analysis methods (c) extract relevant information from the calculated features. The proposed heuristic based-method for quantifying densities and flow(3), is applied to estimate the flow at locations according to parameters derived from the features. Furthermore, the configurable traffic flow analysis is applied to analyse traffic flow between specific areas of the hospital, with traffic classified according to the derived features.

The visualization tools (d) provide intuitive and interactive access to the information extracted in order to facilitate assessment and planning regarding facilities and services in the building complex. The visualization tools show different outputs provided along the entire process as heat-maps, flow-maps, graphs and tables and thereby provide an important set of information that reflects different aspects of the utilization of the buildings, and of associated facilities and services.

5. Feature Calculation

This section covers the proposed rich set of features calculated to enable the mentioned analysis methods. Furthermore, to argue for the feasibility of using large-scale WiFi traces for facility planning we provide illustrating examples of the feature data calculated from the hospital data set.

5.1. Large-scale WiFi Positioning

To estimate the position of the observed mobile devices we use a WiFi positioning module. Since we do network-based measurement collection we will only be able to position devices when they scan for networks. Musa et al. [7] provide statistics and observations of the scanning behavior of different mobile devices, e.g., most devices scan when the screen is turned on or when they aim to transmit data. In the collected data, the median and average time between a device's scans are 58 and 196 seconds respectively, with large variations, e.g. a device may scan every two seconds when active, while it may not scan for half an hour when inactive. Whenever an AP observes a scan it sends to a central machine a measurement message which contains: the id of the AP, the MAC address of the device, the received signal strength (RSS) in dBm, and a timestamp. This is enabled by employing a status-surveillance feature which is common in modern WiFi infrastructures. When enabled, each AP will whenever it receives a message from a device send a TZSP-packet containing the collected information to the central server. The main advantage of using this network-based measuring approach is that every device providing WiFi connectivity can be monitored, independently of its platform and installed software, thus reducing the system deployment time and cost and not requiring the user to install specific software [22]. In order to only track mobile devices, and not infrastructure devices, we filter measurements based on the vendor-specific first three octets of the MAC adresses, as further described in [14].

At the central machine MAC addresses are encrypted and position estimates are computed from the RSS measurements using the centroid lateration algorithm as described in [23]. For this computation the algorithm only requires the location of the APs. Using these, the algorithm estimates the position of a device to be the weighted geometric average of the locations of the receiving APs, using as weights the received signal strengths for each AP. The estimate is then snapped to the location of the nearest AP, in order to enforce that reported positions are inside the buildings. Using this approach the position estimates were evaluated to have a mean accuracy of $15m$ on traces collected through-out the buildings. While other methods may provide more accurate estimates, such as fingerprinting based methods [24], they have additional requirements such as collection of fingerprints or the availability of digital building models. Reliable fingerprint collection (and keeping it up-to-date over time, facing also building- and WiFi-infrastructure changes) at a hospital with more than six thousand rooms spread over ten hectares was deemed unfeasible [22]; furthermore, a complete digital building model, suitable for fingerprinting, of the hospital was not available.

5.2. Classification of Beyond Building Perimeter

Discriminating whether a device is inside or outside one of the complex's buildings is a difficult task as such complexes often have court yards and passages between buildings. Previous work has considered this problem using GPS signals [25] and other sensor modalities [26]. However, given only WiFi measurements these solutions do not apply, and the WiFi positioning literature has also not yet addressed the problem.

In general, when being located outside but close to a building, the WiFi signals emitted from a device can be observed by the APs within the building; a positioning module as described above would therefore end up placing the observed device inside the building. These situations generate erroneous cases in which the device could receive certain information, e.g., from an indoor navigation application or advertising from a specific shop, when it is still out of the buildings that offer these services. In the chosen scenario such errors may impair our analysis methods, e.g., for detecting the time of entry in a building. Moreover, distinguishing outdoor from indoor positions may allow us to filter out those devices that never enter the building and therefore should not be taken into account in statistics of people utilizing the building facilities. We employ a distinction algorithm which uses machine learning based on features extracted from the signal strength measurements, specifically: (i) The signal strength difference between the strongest AP and the weakest AP observed; (ii) Average signal strength of the k-strongest APs received; (iii) Averaged distance between the device's estimated position and the position of the k-strongest APs received; (iv) Average distance among all the received APs; (v) Percentage of *perimeter APs* observed: for this we define a perimeter area utilizing the building complex's layout data, and we classify a AP as either a perimeter AP or interior AP according to if it is within or out of the perimeter area. The perimeter area covers those APs that are within a fixed distance of the actual perimeter. In Figure

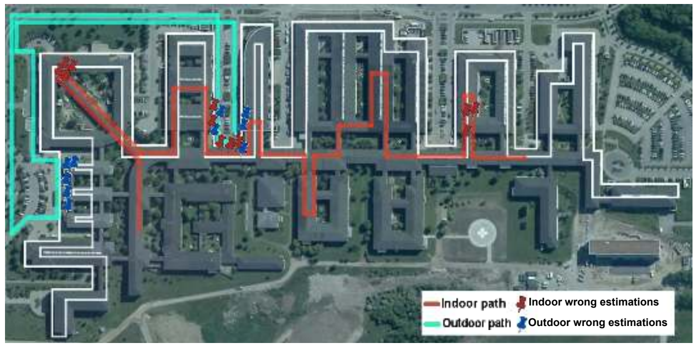


Figure 3: Geometry that defines the perimeter area of the building complex. Indoor and outdoor paths and examples of wrong estimation cases.

3 the perimeter area is highlighted as it is defined for the hospital; note, that this area includes only the part of the perimeter that is physically accessible from public streets. In [14] we provide the algorithm, its implemenation and evaluation in detail. The overall accuracy we perceived in our experiments was at around 95 percent.

We are conscious that the features listed above may need to be adjusted in order to use the classifier at other building complexes according to their wireless network infrastructures. For instance, for high-rise buildings also the floor level detected by the positioning system can provide valuable input to the classifying procedure. Furthermore, our analysis detailed in [14] revealed that among the listed features the ones having the strongest benefit for the intended classification are features iv) and v); these features are largely independent from specific device's hardware characteristics (e.g. from absolute RSSI value computations), and thus can cope well with device heterogeneity [27].

In Figure 4, incorrect classifications are labeled by colored pushpins: red when a pedestrian walking indoors was classified as outdoors and blue when a pedestrian walking outdoors was classified as indoors. Note that the location of the pin is that which is estimated by the positioning system independently of the classification. As one can observe, most of the incorrect classifications in both test cases, indoors and outdoors, happen near main entrances or in areas where the number of APs detected is relatively low (which is the case in one of the building corners). Although entrances are a crucial challenge when distinguishing inside and outside positions, the problem can be alleviated since people are not constantly leaving and entering a building in short order: i.e, when they are inside or outside the building, they usually remain so for a sufficiently long time to produce several position estimates. This in turn allows us to optimize the robustness of the estimations, c.f. Section 6.1.

5.3. Calculation of Features

A crucial task for the goal followed in this paper and when dealing with large sets of unlabeled data is the design of features for extracting vital information on which further analysis can build. In the following we list features central to this task, differentiating them into three categories: temporal, spatial and spatio-temporal.

Temporal features capture aspects concerning the times a device is located within the building complex.

Number of days detected (T1) indicates the number of days we observe a specific device, as shown in Figure 4a. Within the chosen use-scenario the feature helps distinguishing between devices that belong to employees and those that belong to visitors, since the duration of observations should be clearly different in those cases.

Hours per day (T2) spent inside the building complex. In general terms, employees' smartphones remain visible within the building more hours per day than those of short-term visitors, but less than those of hospitalized persons. Such differences can be observed in Figure 4d, where *Device 1* is typical for a short-term visitor whereas *Device 4* is typical for a hospitalized person.

Daytime (T3) indicates the times of day each device is observed. We distinguish between: during day-time(e.g. 7am to 11:59pm), night-time (e.g. 9pm to 6:59pm) and during both. As shown in Figure 4d, devices of hospitalized persons (*Device 4*) are usually observed at any time, whereas visitors are mainly observed during daytime.

Working shifts (T4) help us to discriminate what devices belong to employees or other people that have a fixed timetable. Since the ranges of working hours can vary from one environment to another, we have taken into account the hospital working shift schedule (from 7am to 3pm, from 3pm to 11pm, and from 11pm to 7am). Figure 4c shows the number of devices whose duration inside the hospital correlates with a shift time on at least three days. Those devices would clearly belong to employees.

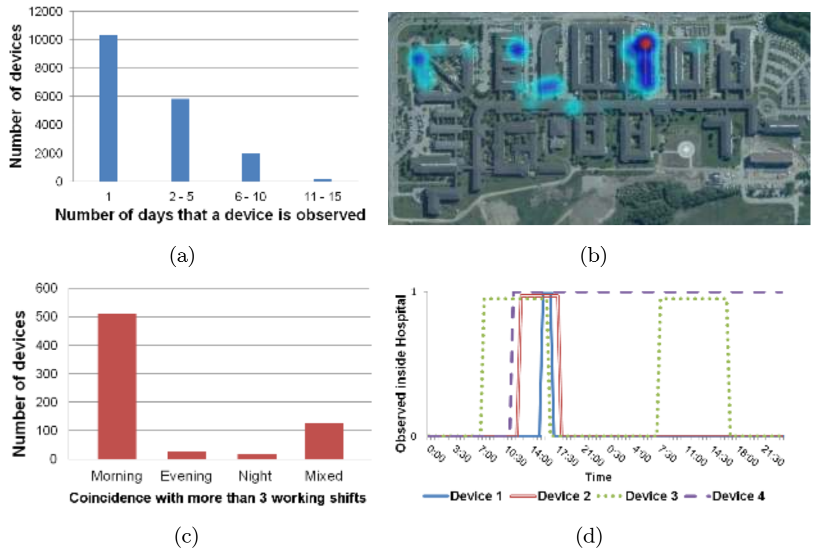


Figure 4: (a) Number of unique devices grouped according to the number of days they were observed. (b)Areas where a device spent most time stationary. (c)Statistics about working shifts.(d) Time of detection inside the hospital of four different devices representing the four expected behaviors.

Spatial features capture aspects of the locations of people (respectively their devices) within the building complex.

Restricted areas (S1) indicates that a device resides within hospital areas that are restricted to certain kinds of people; for example, surgery rooms and laboratories. The areas accessible only to employees are indicated in Figure 1. Moreover, in the particular hospital most parts of the basement floors are only accessible to employees. This last restriction, *times observed in basement* is one of the features that will be used in the posterior processing.

Frequent places(S2) determines the set of areas were a device is frequently observed. This information allows, e.g., to infer ambulant treatment types or job roles.

Beyond Building Perimeter Classification (S3) has been described in Section 5.2 and is listed here for completeness.

Spatio-temporal features consider both spatial and temporal aspects of a device's movement within the building complex.

Motion speed (TS1) depicts average speed of a device. The feature's accuracy depends on realised positioning accuracy as well as frequency. We estimate speed based on the distance covered over time. Though this does not provide a highly accurate speed estimation, it serves well to differentiate motion status (still vs. moving) of devices. Earlier work [28] has proposed a more accurate method for still vs. motion detection using raw signal measurements, however, we did not apply this method because it requires frequent measurements often not satisfied in our data set.

Time stationary (TS2) reflects whether a device has been stationary for a longer period of time—which we define here as being located for more than T minutes within r meters of any single place. For choosing r we suggest taking into account the average distance among APs.

Places where stationary (TS3) determines, relating to the feature S2, the different locations where a device has been stationary, e.g., in a waiting, patient or meeting room. Figure 4b shows a building map indicating the places visited by a device during one day (with the color scale indicating total stationary time at the respective locations).

The presented features form the basis for the analysis methods presented in Section 6. Furthermore, the graphical presentation of the collected data set for the described features illustrate and highlight their utilization, revealing e.g., that ca. 2000-3000 mobile devices were observed per day, and that a large fraction of these were observed only on one day (Figure 4a). These numbers support that our measurement approach provides rich data for a significant number of devices. Compared to previous wireless network studies in campus or company settings [8, 9, 10], the large percentage of one-day-only visitors differentiates this data set from what has been observed in the above studies where the sets of perceived devices per day were highly correlated across days. This also highlights that hospital environments are different and thus relevant use-scenarios to consider in future work in wireless network analysis and related fields.

6. Analysis Methods

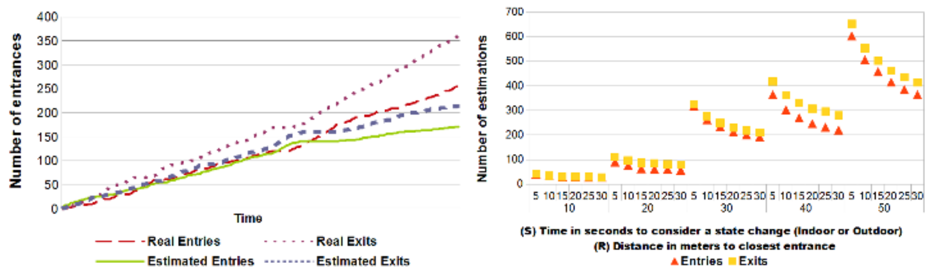
In the following, we describe how to utilize the features extracted from WiFi measurements for further analysis methods for informing and supporting decisions within facility utilization analysis, focusing on aspects introduced in Section 4.

6.1. Density and Flow Estimation

The density of people in a specific area or the flow of people through a given area or across a given line or other borders are fundamental types of information within planning in both indoor and urban settings [11]. To obtain such information, we propose to apply a number of heuristics using the features introduced in Section 5. In the following, we will consider the specific case of quantifying the flow through entrances as people enter and leave the hospital. Such information enable the deduction of, e.g., the most used entrances to a building complex, which helps to decide e.g., where to install information boards or vending machines (since these would be the most busy areas), or to determine the flow-wise most appropriate entrances for emergency cases (i.e.,less crowded ones), or to determine where to build additional parking places (i.e., in those areas by which people usually enter into the hospital), or to design evacuation plans (for individual day-times or weekdays, or even dynamically, according to current crowd conditions, among others).

To estimate the flow through entrances we propose a method building on the beyond building perimeter classification from Section 5.2. Having calculated the beyond building perimeter feature value, once we detect a change in the device's in/outdoor state, we record its timestamp. To avoid erroneous rapid state changes provoked by signal variability in devices which scan frequently, the method waits for the new state to remain stable for at least S seconds before it registers a new entry or exit event. We assign the event to the closest entrance among a list of entrances previously defined. To avoid false positive cases we record the event only in case the distance between the closest entrance and the estimated position is below a threshold R.

To evaluate the method's accuracy, we have carried out several empirical tests using different configurations for the threshold parameters S and R which define whether an entry/exit event should be recorded. Figure 5a depicts the number of entry and exit events that have been estimated at the hospital's main entrance over a period of 6 hours. During this time, a person manually counted the number of actual entries (327) and exits (453) at the entrance, obviously with no knowledge about how many of people that were counted also carried a smartphone. We can assume that the ratio of smartphone holders is close to the 59% reported as the estimated percentage



(a) (b)

Figure 5: (a) Entry and exit events over time for $S\!=\!30s$ and $R\!=\!40m$. (b) Total entry and exit events for different S and R values.

(a) (b) (c) (d)

Figure 6: (a) Heat-maps representing all device positions; (b) Only inside to outside movements (leaving the building); (c) Positions of filtered exits; (d) Estimated exits constrained to real exits.

of smartphone penetration in 2013 in Denmark[1]. These numbers would correspond to 192 entries and 270 exits of persons with smartphones approximately. Using these values we calibrate the parameters R and S as shown in figure 5b, and find that the optimal values are $S\!=\!30$ seconds and $R\!=\!40$ meters. Using these configuration values we are able to approximate the expected results as shown in Figure 5a. Finally, in order to provide visualizations of various obtained results on the complete data set we build, after executing the method, heat-maps as shown in Figure 6.

6.2. Configurable Traffic Flow Analysis

In addition to flow at specific locations, we propose methods for analysis of the flow of traffic between configurable areas of the covered buildings. The methods are designed to be configurable in order to allow for input from e.g. domain expert users who wish to perform analysis for specific buildings or departments, a specific time frame, or for specific classes of building users.

The areas of the hospital between which traffic flow should be analysed are specified as polygons whose spatial extend each covers an area of interest. These polygons are specified in the open KML format, which allows for easy configuration of the areas using editors such as Google Earth. During the traffic flow analysis computation, each position estimate is annotated with the area of the polygon it is contained in, if any. After this annotation with areas, the position estimates are aggregated into areas, so that several consecutive position estimates in a single area defines a single time-interval within that area. From this information we can easily extract the number of travels between the various areas. Optionally, the travel data may be filtered w.r.t.: Minimum and maximum time in an area as well as minimum and maximum time between areas. These filters may be used to e.g. ignore stays in an area of less than 5 minutes if areas which are just passed through are not interesting for the analysis. Note here that not all devices perform WiFi transmissions at frequent intervals, which may influence the analysis results, as some traversed areas may not be registered. However, a frequency analysis of the collected data shows that 80% of devices allow for position computations at least every 5 minutes. This allows for registering any areas which are visited for more than a short period, but areas which are simply passed through, such as the hallways, may be underrepresented in the data. The maximum time between areas can be used to avoid the situation of registering e.g. an employee going home and coming back the next day as a travel between departments. In addition to the filtering, the travels may be classified according to the features defined in Section 5, e.g. in order to analyze traffic flow based on work shifts.

7. Visual Analysis

In this section, we present and visualize tools built based on the introduced features and analysis methods. We furthermore present respective results for the test-bed hospital, and discuss their use for facility utilization assessment and improvement.

7.1. Visualizing Traffic Flow

To illustrate how visualizations based on WiFi monitoring and our analysis methods are generated, Figure 7 shows individual movements observed from a hospital entrance until the device reaches a stationary destination (waiting room, canteen, office, etc.) as described by the "places where stationary" feature (TS3).

Figure 7: Tracking three devices from a particular to their respective destination.

For privacy reasons, these visualizations are computed from traces collected by the authors. The left part of the figure shows the real paths while the right one reflects the estimated paths. The obtained results support that our method is valid as in all three cases the correct entrances and stationary end point was detected by our methods.

The concrete entrance chosen here has a high load, see Figure 6d—higher than intended given its location; noteworthy is also that the closest main entrance has a comparatively low load. For further analysis, it is relevant to consider where people using an entrance end up within the building. Given, e.g., the obtained 15 day data set, our methods aid in such analysis and in assessing if the the paths people currently take are optimal, or whether instead means for improved directing of flow would yield improved efficiency or safety.

7.2. Travel-graph Based Analysis

Generalizing from the individual movement data discussed above, we now present visualizations of aggregated traffic flow between locations within the test-bed hospital. Domain experts advised us on their needs for visualizations and analysis tools, and the produced visualizations were evaluated in collaboration with them. We interviewed two hospital professionals, respectively a project manager from the hospital planning department, and the head of a logistics department at a large hospital. The evaluations were performed through a semi-structured interview, where the hospital employees were presented with printed versions of the various visualizations. They were asked about the correctness of the traffic flow as well as usefulness of the visualizations, while notes were taken by the interviewer. Figure 8 and 9 show traffic flow, computed and filtered as described in Section 6.2, during the complete observation period and for all detected mobile devices on three different spatial resolutions; in each graph the edges reflect travels between locations, where the set of locations are: individual locations in Figure 9a; individual buildings in Figure 8 and department complexes in Figure 9b, as per Figure 1. In each graph the thickness as well as the color of edges encodes the number of recorded travels, according to the scale given with each graph, from lowest (thin green) to highest (thick red). As edges are intended to show completed travel, the latter needs to be defined; for the graphs shown a travel end is constituted by the device being observed for longer than 10 seconds. Such a filter aids in removing also spurious inaccurate position estimates to neighbouring areas. The interval may be increased depending on the analysis, e.g. it may be interesting to note the wards between which people travel and then stay for longer periods of time.

Combined, these graphs allow to investigate the facility utilization and whether optimizations, e.g. in the distribution of facilities are required. The three different granularities facilitate different aspects of, and interests within, facility utilization investigation: The fine granularity graph in Figure 9a allows to identify visually, e.g. the most common routes taken, as well as traffic bottleneck zones, and aids in finding potential solutions for these, e.g. in the form of additional pathways or alternative signs for guiding visitor traffic. The two other graphs, showing travels between buildings and departments, identifies which buildings and departments have the highest interaction with each other, and thus should be ideally i) spatially close and ii) well connected by pathways. We chose this form of visualization of the data as it is easy to determine the spatial relation between buildings, and thus to check whether large flows between departments are simply due to spatial closeness, or whether it may be caused by other factors. However, the data may be visualized in other ways as appropiate, e.g. the project manager suggested that a matrix enumerating the flow between different wards would be very beneficial for layout-planning purposes, instead of the current process of asking individuals at each wards with whom they collaborate. For the following exemplary illustration of travel-graph-based analysis, we focus on the building-level granularity, as shown in Figure 8 where buildings are labelled by letters, and in further graphs which visualize selective portions of the recorded travel data on building-level granularity. As we do not have full ground truth due to the large scale of the hospital, we have instead had the two hospital professionals evaluate whether the detected movement

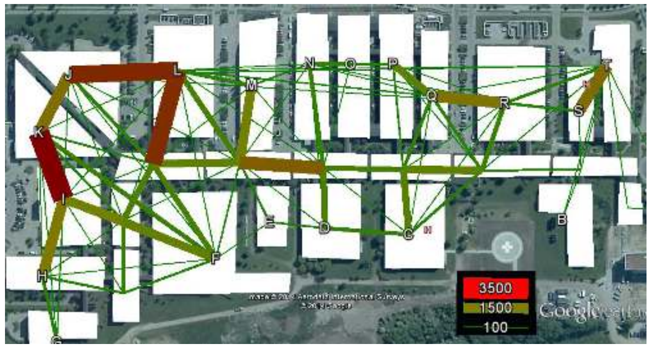
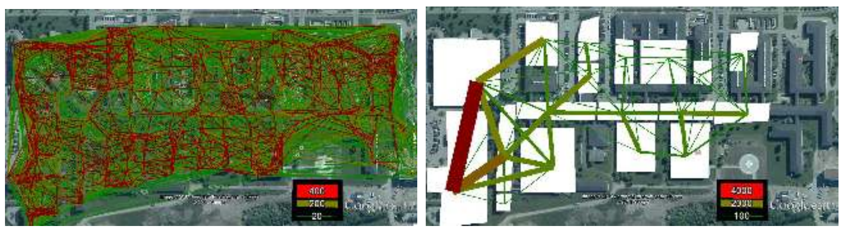


Figure 8: Overview of movement between departments as well as hallways.

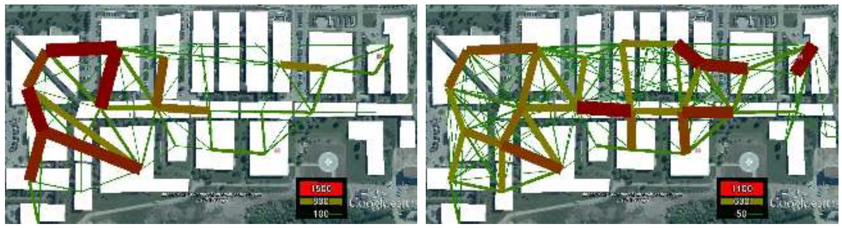


(a) (b)

Figure 9: Overview of movement at two levels of detail: between individual locations (9a) and between specific departments (9b)

patterns are supported by their professional knowledge of the activity at the hospital.

Detailing for individual user roles. The recorded travel data, as shown in Figure 8 and 9, can be divided by user roles, into data for visitors, ambulant patients, or staff, respectively, as they can be inferred, e.g., by employing behavioral classification of users as described in [14]. Figure 10 allows for this comparison, and makes for several interesting observations. The graphs suggest that while the majority of guests seem to move around the left area of the hospital, the employees travel throughout the entire hospital and hallways. According to the hospital professionals, the left area of the hospital contains the maternity wards and pediatric wards, which have a large flow of people due to pregnancy-checkups, visits to new mothers, and parents visiting hospitalized children. The right area of the hospital on the other hand contains mostly wards, where patients stay for several days at a time, and laboratories to which visitors do not have access. From a facility-management point of view, the figures indicate that visitors are largely constrained to one part of the hospital—yielding the benefits that i) visitors have only a small and less complex area to travel and to be guided in, and that



(a) (b)

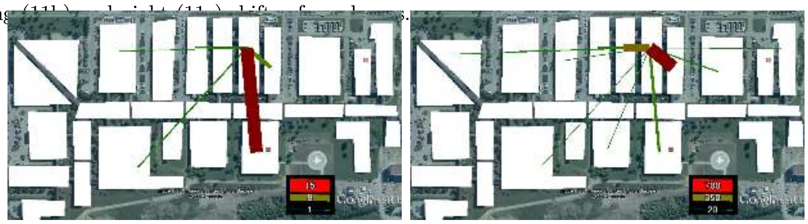
Figure 10: Overview of movement separated into that of visitors (10a) and employees (10b).

ii) visitors do not interfere with e.g. work at the surgery wards where unobstructed movement along the pathways and also a more quiet environment are desired.

Detailing for daytimes. The travel data, as shown in Figure 8 and 9 can be divided temporally, e.g., in week- vs. weekend-days, or according to the shift period they fall into. Figure 11 shows building-level travels during day, evening and night shifts, respectively, as defined in feature T4 in Section 5. Comparison of the three figures reveals as expected the highest level of activity during the day shift, slightly less activity on evenings, and a very quiet hospital at night, when primarily only strictly necessary or emergency tasks are performed. The comparison shows also a change in the spatial patterns of activity: E.g., the building hosting the personnel-cafeteria (M) is traveled to and from often during day-time from many buildings, while travel activity in this building is much less distinct on evenings and almost disappears at night when the cafeteria is closed. Similarly, several wards see a lot less activity during evening hours, e.g. building F, D and C at the lower mid of the hospital, as the consultations hosted in these are scheduled primarily during daytime. The hospital professionals further pointed out that the decrease in activity at evening and night time around building S and T are due to those primarily being laboratories in which mostly daytime work is performed. Figure 11c furthermore indicates that during nighttime the building most frequently visited by employees is C, which hosts emergency reception for patients with acute heart issues. This again suggests a beneficial layout of buildings in regards to facility-management, as the night employee activity is centered on a very small part of the hospital, thus avoiding i) the disturbance of e.g. hospitalized patients, and ii) unnecessary long traveling between tasks.

(a) (b) (c)

Figure 11: The number of travels between hospital departments and hallways, during day (11a), evening (11b) and night (11c) shifts [illegible]s.



(a) (b)

Figure 12: The number of travels to and from the surgery ward for patients (12a) and employees (12b).

Detailing for spatial focus areas. The methods also allow for focusing on specific places of interest. Figure 12 shows the movement, for patients and employees, respectively, to and from a specific building containing parts of the surgical ward as well as performing some outpatient treatment. It shows that the building which supplies the most patients for surgery or outpatient treatment is the one directly below, which contains consulting rooms as well as the emergency reception for heart issues. The hospital professionals confirmed that patients often are moved from building C to P for surgeries, but that they are generally not allowed to carry smartphones, only when heading for the outpatient clinic located there. This may be why the numbers for patients are so much smaller than those for employees. They also mentioned that the buildings around P, as well as C, all share the same staff, which moves around these buildings regularly. This could be why we see a lot of movement between the neighbouring buildings and C in the figure for employees. Of note is also the movement to and from building R which contains the blood bank, and as such is likely due to employees transporting blood for surgeries.

While the hospital professionals were overall positiove regarding the correctness of the results, it's worth mentioning the unexpected results. They noted that maternity wards ought to have more activity at night, as this is naturally busy through all 24 hours. Additionally, there is an unexpected high amount of movement directly between J and L. This may be due to errors in the position estimates when persons close to the perimeter of one of the buildings are falsely located in the other.

8. Conclusions

In this paper we have proposed a rich set of features and analysis methods to inform building facility planning enabling studies of people's behavior in large building complexes utilizing solely measurements of WiFi signals from peoples' devices. To this end, we have addressed the challenges coming with the complexity of the chosen environment. To the best of our knowledge, this is the first study of its type which addresses hospital complexes. The proposed analysis methods include a method to estimate when and where users (respectively their mobile devices) enter and leave buildings. This addresses shortcoming usually inherent in the WiFi-based tracking and offers several possibilities, e.g., to analyze the flow of people from the specific moment they enter a building. In addition we provide methods and visualization tools for analysis of the traffic flow between specific areas of the hospital, according to features such as user group or time of day. This further empowers the facility-management by enabling domain experts to perform specific analysis to determine whether the facilities are utilized optimally. We achieved the central goal of providing realistic information that reflects realistically the behavior of, e.g., hospital staff or visitors who make use of the facilities and services offered. Thus, the proposed methods can provide valuable sources of information, e.g. regarding building, path and service utilization, for supporting hospital planning activities.

Building on presented results, for future work we plan to evaluate analysis methods for further aspects of human behavior and consider the development of privacy protecting methods to enable gathering of labeled data in hospital environments.

Acknowledgments

The authors acknowledge the support granted by the *Danish Advanced Technology Foundation* under J.nr. 076-2011-3. Antonio J. Ruiz-Ruiz was supported by the Seneca Foundation under the Seneca Program 2009, the Spanish MINECO, as well as EC FEDER funds, under grant TIN2012-38341-C04-03.

References

[1] N. Edwards and A. Harrison, "The hospital of the future: planning hospitals with limited evidence. a research and policy problem," *British Medical Journal, 319: 1361*, 1999.

[2] R. B. Bachouch, A. Guinet, and S. Hajri-Gabouj, "An integer linear model for hospital bed planning," *International Journal of Production Economics*, vol. 140, no. 2, pp. 833 – 843, 2012.

[3] G. Ma and E. Demeulemeester, "A multilevel integrative approach to hospital case mix and capacity planning," *Computers and Operations Research*, vol. 40, no. 9, pp. 2198 – 2207, 2013.

[4] P. VanBerkel and J. Blake, "A comprehensive simulation for wait time reduction and capacity planning applied in general surgery," *Health Care Management Science*, vol. 10, no. 4, pp. 373–385, 2007.

[5] A. Marshall, C. Vasilakis, and E. El-Darzi, "Length of stay-based patient flow models: Recent developments and future directions," *Health Care Management Science*, vol. 8, no. 3, pp. 213–220, 2005.

[6] B. Rechel, S. Wright, J. Barlow, and M. McKee, "Hospital capacity planning: from measuring stocks to modelling flows," *Bulletin of the World Health Organization*, 2010.

[7] A. B. M. Musa and J. Eriksson, "Tracking unmodified smartphones using wi-fi monitors," in *Proc. of SenSys.* ACM, 2012, pp. 281–294.

[8] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. of MobiCom.* ACM, 2004, pp. 187–201.

[9] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. of MobiSys.* ACM, 2003, pp. 303–316.

[10] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Usage patterns in an urban wifi network," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1359–1372, 2010.

[11] E. O'Neill, V. Kostakos, and T. e. a. Kindberg, "Instrumenting the city: Developing methods for observing and understanding the digital cityscape," in *Proc. of UbiComp.* ACM, 2006, pp. 315–332.

[12] A. Millonig and G. Gartner, "Identifying motion and interest patterns of shoppers for developing personalised wayfinding tools," *J. Location Based Services*, vol. 5, no. 1, pp. 3–21, 2011.

[13] B. E. Moore, S. Ali, R. Mehran, and M. Shah, "Visual crowd surveillance through a hydrodynamics lens," *Commun. ACM*, vol. 54, no. 12, pp. 64–73, 2011.

[14] A. J. Ruiz-Ruiz, H. Blunck, T. S. Prentow, A. Stisen, and M. B. Kjaergaard, "Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on.* IEEE, 2014, pp. 130–138.

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

[16] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, "Automatically characterizing places with opportunistic crowdsensing using smartphones," in *Proc. of UbiComp.* ACM, 2012, pp. 481–490.

[17] L. Vu, Q. Do, and K. Nahrstedt, "Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace," in *Proc. of IEEE PerCom*, 2011, pp. 54–62.

[18] M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster, "Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones," in *Proc. of UbiComp.* ACM, 2012, pp. 240–249.

[19] M. B. Kjærgaard and H. Blunck, "Tool support for detection and analysis of following and leadership behavior of pedestrians from mobile sensing data," *Pervasive and Mobile Computing*, vol. 10, Part A, pp. 104 – 117, 2014.

[20] M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster, "Mobile Sensing of Pedestrian Flocks in Indoor Environments using WiFi Signals," in *Proc. of IEEE PerCom*, 2012.

[21] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. M. T. Do, O. Dousse, J. Eberle, and M. Miettinen, "From big smartphone data to worldwide research: The mobile data challenge," *Pervasive and Mobile Computing*, vol. 9, no. 6, pp. 752 – 771, 2013.

[22] M. B. Kjærgaard, M. V. Krarup, A. Stisen, T. S. Prentow, H. Blunck, K. Grønbæk, and C. S. Jensen, "Indoor positioning using wi-fi–how well is the problem understood?" in *International Conference on Indoor Positioning and Indoor Navigation(IPIN*, 2013.

[23] A. LaMarca, Y. Chawathe, S. Consolvo, and J. H. et al., "Place lab: Device positioning using radio beacons in the wild," in *Proc of IEEE PerCom*, 2005, pp. 116–133.

[24] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *Proc of IEEE InfoCom*, 2000, pp. 775–784.

[25] H. Blunck, M. B. Kjærgaard, and T. S. Toftegaard, "Sensing and classifying impairments of gps reception on mobile devices," in *Pervasive*. Springer, 2011, pp. 350–367.

[26] P. Zhou, Y. Zheng, Z. Li, M. Li, and G. Shen, "Iodetector: a generic service for indoor outdoor detection," in *SenSys*. ACM, 2012, pp. 113–126.

[27] M. B. Kjærgaard and C. V. Munk, "Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength," in *Proc. of IEEE PerCom*, 2008, pp. 110–116.

[28] T. King and M. B. Kjærgaard, "Composcan: adaptive scanning for efficient concurrent communications and positioning with 802.11," in *Proc. of MobiSys*. ACM, 2008, pp. 67–80.