Implementation of a virtual occupancy sensor for smart building support

J. Aizpuru, R. Jentsch, M. Šipetić, G. Zucker

AIT Austrian Institute of Technology, Austria

Abstract

Advancements and wide availability of cheap sensors are fuelling the development of Internet of Things (IoT) applications. More sensors mean more data, and more relationships between data points need to be tracked in order to effectively understand, manage and control systems. Having access to either real-time presence data or patterns extracted from historical data is particularly valuable when dealing with facility design and management due to their direct correlation with energy consumption and indoor comfort. We propose a straightforward, cost effective and privacy-preserving method to extract the occupancy information. By aggregating semantic knowledge, motion sensor data and data from dwelling entrance doors, a robust virtual occupancy sensor has been developed; it is underpinned by an ontology that was developed on top of the set of standard ontologies like Building topology ontology (BOT) and Smart appliance reference ontology (Saref) that allowed describing all relevant datapoint and demo site metadata and enabled automated processing of collected data. The method is replicable to all built environment described in a similar way where motion information is collected and where there are clear boundaries of monitored space, and the occupancy information can be useful in different application cases. While predictive occupancy models or expensive sensing alternatives have been already exploited for similar purposes, our solution is simple, inexpensive, replicable and easy to implement in existing buildings.

Keywords: facility management, smart building, semantic web, occupancy, residential, accommodation

1. Introduction

The building sector and activities of people in buildings are responsible for approximately 31% of global energy demand (Ürge-Vorsatz et al., 2012). Thermal comfort services (heating, cooling, ventilation) account for 36% of operational energy demand in buildings. Depending on the region and building sector, these loads are usually controlled either manually or using predefined schedules. While predefined schedules are effective in reducing unnecessary energy usage, once they are commissioned, they are usually not adapted to changing conditions. Operation of thermal comfort services based on occupancy has been proposed before and has been shown to offer significant savings compared to fixed schedules and manual control.

Occupancy detection is typically implemented by using infrared motion detectors that signal the change of temperature. Mathematically it is a nonlinear integration over the temperature in the field of sight. These sensors are affordable and are therefore widely deployed in buildings, typically in combination with a timer that deactivates after a configurable time period. The shortcomings of these sensors are twofold: first, they are not presence detectors i.e. a person that does not move, is not detected; second, persons are only visible if they have a temperature difference to the environment, thus the accuracy decreases in hot environment or when people are fully covered in clothing (e.g. outside in winter season). Still, this type of sensor is the most common solution for applications like automatic hallway lighting, sanitary facilities and generally rooms where people are moving in and out.

Optical sensing methods improve the quality of presence detection but have the problem of compromising privacy; feasible solution is an edge device that pre-process the data and reduce information to anonymous data like a person count. Such devices are available on the market but serve a higher price segment than is commonly feasible for residential or office building applications.

The authors in (Zucker et al., 2017) followed another approach that does not employ any additional sensors: by using the available sensory equipment in the ventilation system, it is possible to make a good estimate of the number of people in a room. The method requires to configure the volume of the room and an estimation of the average CO_2 production of a person. It then calculates the occupancy by using the CO_2 sensor in the ventilation system together with the supply and exhaust air flow. Such approaches are used for advanced heating, ventilation and air conditioning (HVAC) control algorithms like the dynamic reset of volume flow (Lin and Lau, 2014). A similar approach was done in (Mir 2018) where the CO_2 level was fed into a proportional-integral-derivative (PID) controller that minimized the estimation error of occupancy dynamically. Authors in (Yang et al., 2013) compared performance of different machine learning techniques for binary and multi-class occupancy estimations using varying combinations of commonly used indoor sensors, and concluded that CO_2 sensors, door status and light sensors contributed the most to the accuracy of occupancy detection.

In the present work, a simple, cost-effective and privacy-preserving methodology for occupancy estimation has been developed. Making use of a semantic knowledge base which contains all the relevant information about an apartment building topology and device location, linked with a database of timeseries data from the appliances, we have generated a rule-based algorithm which infers occupancy state and uses it as a feature for consumption prediction models.

Concerning the semantic knowledge base, resource description framework (RDF) is a widely used standard for information exchange between different sources (*RDF - Semantic Web Standards*, n.d.). Atom of information in RDF is a triple consisting of **subject**, **predicate** and an **object** (Fig. **1**). This simple structure allows one to express relationships (predicate) that two concepts (subject and object) are in.

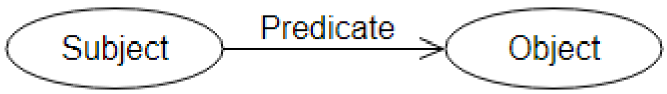

Fig. 1: Graphical representation of a triple

A collection of triples described using RDF forms a directed labelled graph. This enables creating complex network of relationships among concepts. Resulting graph can be explored (traversed) by each application needing the data contained therein, as long as different applications have a common understanding what nodes and relationships in the graph mean. Ontologies are used to define standard concepts and relationships that should be used by all complying application in order to have the same understanding of concepts used in the information exchange. In addition to data representation, it is also possible to deduce and infer the information that is not explicitly stored in the knowledge base.

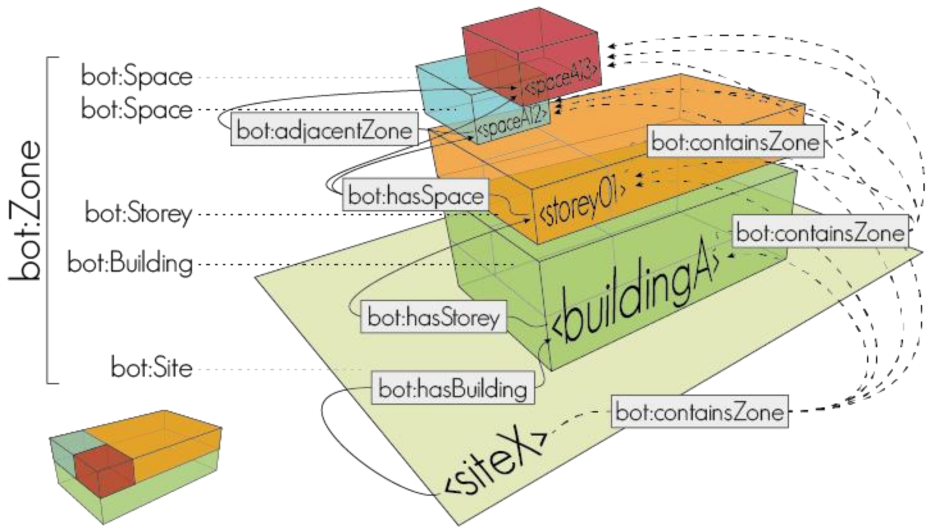

Fig. 2: Representation of spaces and their relationships in a building defined in bot ontology (Rasmussen, Pauwels, Lefrançois, et al, 2017)

In architecture, engineering and construction (AEC) industry several well-known ontologies, such as ifcOWL (Beetz et al., 2009) and Building Topology Ontology (BOT) (Rasmussen et al., 2017), are used to represent the architectural, structural and topological information about the building. ifcOWL can directly represent the information contained in an Industry Foundation Classes (IFC) files as an RDF graph. IFC (*Industry Foundation Classes (IFC)*, 2020) is the standard developed by buildingSMART International and BOT is an ontology focusing entirely on representing topological information about building spaces and elements and is aimed at being a building block for implementation of domain specific extensions on top of it. Fig. 2 shows how concepts defined in a bot ontology are used to represent spaces and their relationships in a building. Similarly, for describing building equipment, several ontologies have been proposed that are in active development. These include Smart appliances reference ontology (SAREF) (Daniele et al., 2015), and its extensions, among which the most relevant is SAREF4BLDG (Poveda-Villalón and García-Castro, 2018) targeting specifically building devices, and Brick ontology (Balaji et al., 2016). For specific use cases described in this work we used a custom ontology (Šipetić et al., 2020) built on top of BOT, SAREF and SAREF4BLDG ontologies.

2. Data

Most relevant information for the developed method for presence detection are detected motion, door opening and closing, and topology of monitored spaces and placement of corresponding sensors therein. As data sources, an InfluxDB database for collecting all the data recorded by the sensors installed in several apartments of a residential building and, the ontology describing the topology of this specific site are available.

2.1. Sensor timeseries

Real time sensor data from a group of apartments are collected in the database. Data collected includes measurements of indoor conditions (temperature, motion, humidity), power consumption and door/window opening information. Each measurement is recorded in the database on-change with a different change threshold for each measurement type, which is advantageous for battery life and storage savings, but may lead to ambiguous information due to the uncertainty, whether the sensor is operational or not when no measurements are recorded for an extended period. To deal with this issue, we have taken advantage of the multi-sensing feature of the devices. Each physical sensor device measures several properties; for example, motion sensors also provide temperature and illuminance measurements, as well as the strength level of the sensors' signal to the gateway. This way, each time the sensor records one of its measurements, it is marked as operational. Missing data can be partially filled using this methodology. For still remaining gaps, different thresholds have been defined for each type of the measurement, which limit the maximum time a sensor can be inactive. After the measurement dataset gaps are filled, resampling is performed on a 15-minute interval to align all time series to the uniform time grid. The method of aggregating the values when resampling depends on the sensors' measurement type:

- Continuous variables – temperature, illuminance, humidity, … – are forward-filled
- Binary variables – motion, door/window opening, … – are summed up

2.2. Semantic knowledge base

In addition to raw time series data from sensors, an essential part of the approach is utilization of semantic information about the location of sensors and the topology of spaces. The demonstration site is completely semantically described, and this description is available in machine-readable form. In this section we present more details about the structure of static information.

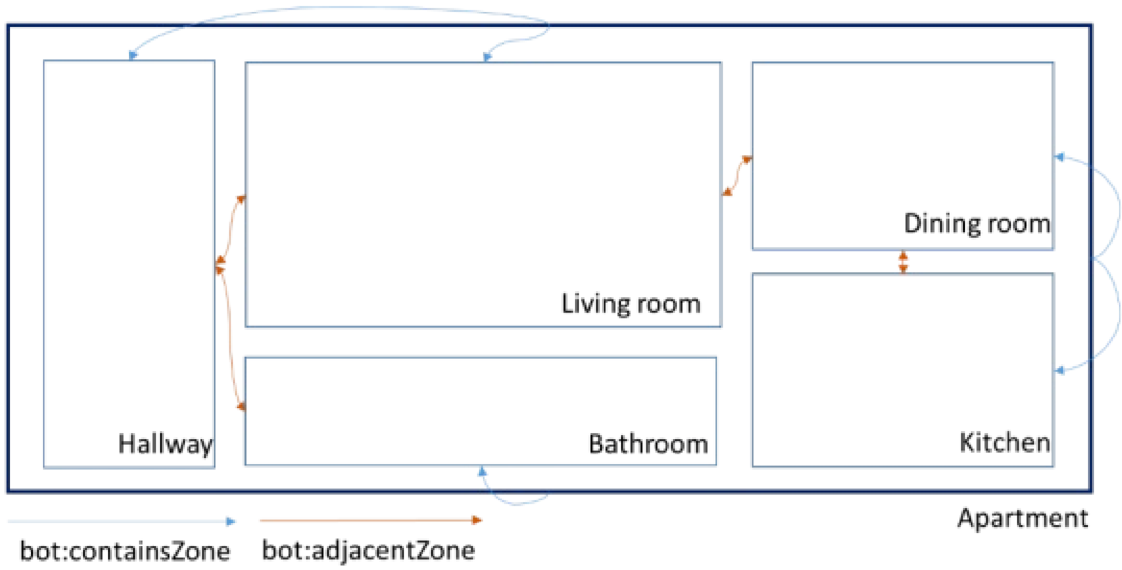

Fig. 3: Topology of an apartment with different spaces and relationships between them following BOT syntax

In Fig. 3 an example of a topology of an apartment is shown with bold largest box being an apartment, and smaller boxes inside representing spaces in the apartment. An apartment is defined as containing a number of different spaces using the **bot:containsZone** relationship (spaces are named zones in BOT ontology). As data is stored in a knowledge base that has a reasoning capability, each apartment space also has an implicit **bot:containedInZone** relationship in the direction of the apartment. That allows one to use either relationship for data querying depending on the convenience and the use case. Axioms defined in the BOT ontology state a number of such inverse relationships. Similarly, rooms have defined adjacencies between them, and adjacency in this context means the ability to move from one room to the other. In this case all the adjacency relationships (**bot:adjacentZone**) are reflexive, meaning when one is defined between two spaces, there is also an implicit relationship in the opposite direction. That allows one to freely form queries that would match entities from either direction.

Each room has a motion sensor placed in it, and that entrance door of the apartment is monitored by a door opening sensor. Door opening sensors, Fig. 4, are physical devices consisting of two parts which are mounted on the door and door frame, opposite of one another when door is closed. When door is opened, movement of the part mounted on the door causes induction in the stationary part, which is transformed into an “Opened” signal, that is sent to the automation system. Once the door closes again, the opposite “Closed” signal is sent.

Fig. 4: Develco door/window sensor

Fig. 5 shows how relationship between an apartment and its entrance are established. An internal wall is instantiated, which is shared by all apartments having entrance door that is part of the same wall. Internal wall is set to be adjacent to a specific space inside an apartment, so that it is clear in which space person enters first through said door.

The internal wall is linked to the door using the **bot:hostsElement** relationships, and door is linked to the sensor device using the same relationship. Another way of placing sensors is shown in the same figure with the example of second sensor that is linked to a containing room using the **bot:containsElement** relationship. Additionally, each of the physical sensor types installed on the site is modelled as a collection of sensors measuring specific physical properties. Possible properties of interest include, but are not limited to pressure, temperature, humidity, open/close state, flow, speed, etc. Each of the door/window sensors described here - apart from sensors for open/close state of the monitored doors - also contain temperature and humidity sensors. Fig. 5 (right) shows how the sensor device is composed of different sensing units. For example, the door sensor that is hosted by a door in Fig. 5 (left) consists of temperature sensing unit and door state sensing unit.

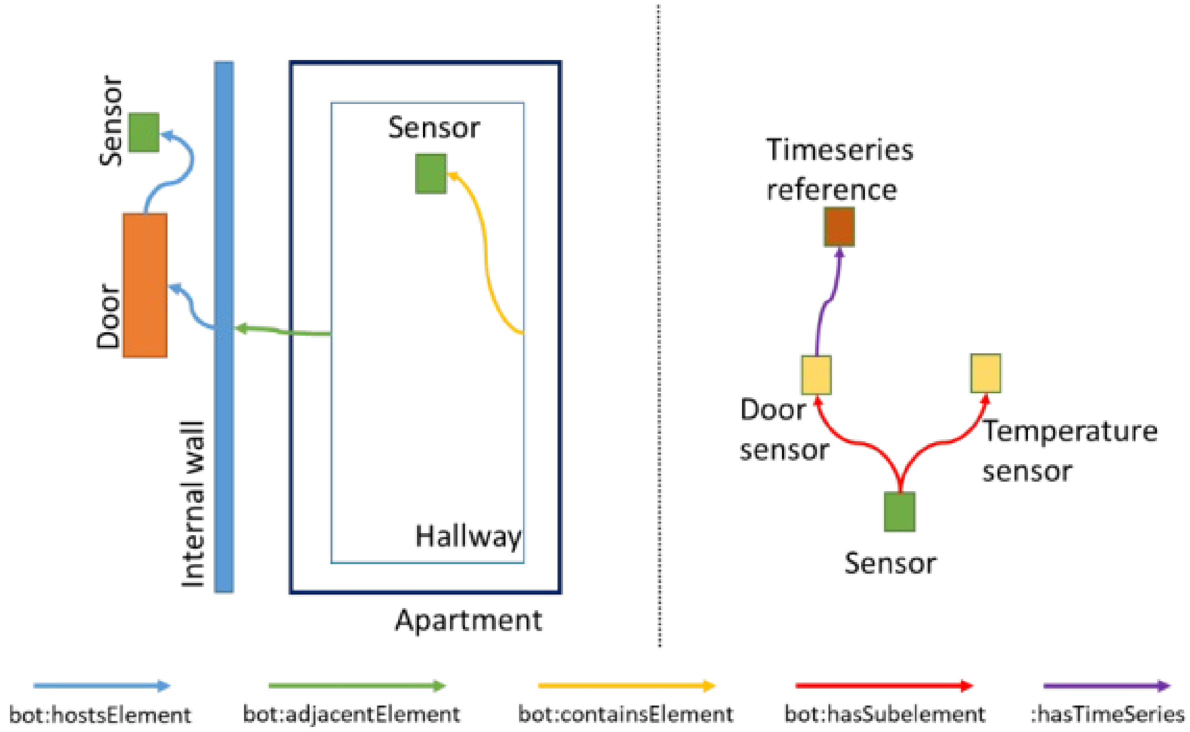

Fig. 5: Relationships established between sensors and building and space elements

Both of these send and store signals in separate time series on the data platform, and reference to these time series is stored in the separate timeseries reference entity that is associated with each sensing unit using the **:hasTimeSeries** relationship. Time series reference contains information such as **database name**, **table name** and **sensor id** when data is stored in the SQL database, or **measurement id** and **field name** when data is stored in time series database such as InfluxDB. Each TimeSeries entity is associated (**:storedInDatabase**) with the containing database as shown in Fig. 6. Database entity describes the specifics (host, ports, …) of each database instance used.

Knowledge bases also allow direct recording of time series into the data model, which allows some interesting use-cases where reasoning is possible over the values of data recorded for each time series. However, for production scenarios, this is usually avoided due to performance reasons, as large amounts of data such as data in IoT scenarios, in combination with reasoning usually cause performance issues.

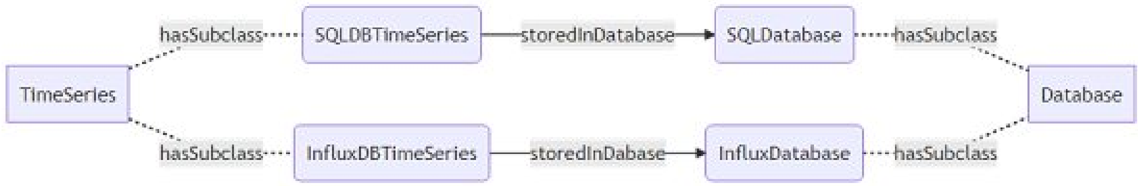

Fig. 6: Metadata structure linking database and time series concepts

3. Methodology

While CO_2 sensor information improves the accuracy of occupancy detection, CO_2 sensors were not available in our demonstration sites. Their high cost prohibits widespread use, especially in residential scenarios. On the other hand, door status and motion data were readily available in project demonstration sites. In addition to that, the space hierarchy, along with sensor and appliance locations and connection information has been collected previously in a separate effort, so all the semantic information was also available. While usage of additional sensors such as temperature, humidity, pressure, smart plugs and similar, could have been possible in some apartments, only the motion sensors were consistently available in all the spaces of the demonstration site, and door status sensors were installed on entrance doors of all the apartments. Limiting the requirements of the technique to just two measurement types simplifies replication. The required sensors are often present in existing smart home installations, and if they are not, retrofit action to add those is simple and inexpensive.

While more detailed space hierarchy, and sensor and device location information are useful in some other scenarios, for the proposed technique we use only the combined information about all the sensors contained in the monitored space, and door status of entrance to the monitored space. If more detailed information is available, it can be easily reduced to this data set, and if the solution is installed from the ground up in the monitored space, the limited data needs reduce complexity of describing the site.

Based on the above, we developed a technique for inferring occupancy using only two measurements: motion recorded within the apartment and entrance door openings. The key for extracting and combining relevant measurements easily from the entire dataset of the building is the knowledge base based on the previously described ontology, which returns the relevant ids of the devices to query from the timeseries database.

To cover different scenarios, two approaches for inferring occupancy will be presented: occupancy feature generation from historic data and real-time occupancy inference.

3.1. Using historic data

This approach is interesting for applications where occupancy information is needed as a historic feature or where the sampling frequency of the dataset is not precise enough, e.g. our use case, where data points are represented in a 15-minute interval.

The initial step is to extract the ids of the apartment's motion and door sensors and filter out all the doors and windows which are not connecting the common areas of the building with the apartment (i.e. all non-entrance door sensors). Window sensors have been filtered out to prevent interpreting state change of an e.g. open window on a windy day as an occupant interacting with the window. The necessary information can be directly queried from the database and processing started. Afterwards, the methodology is straightforward: the data available for inferring historic occupancy is divided into intervals between two door openings-closings and if there is any motion in between, the occupancy state is activated (in binary, 1); in the opposite case, the state is labelled as unoccupied. For the beginning and ending part of the dataset, the state can just be decided if there is any motion and if this information does not exist due to motion sensors not working, the state would be unknown until the next door-motion begins.

3.2. Real-time occupancy inference

For some applications occupancy state information based on real-time processing of data could be useful. In case of streaming operation, it is not possible to process intervals between door openings and closings as they are not available in real time. Instead the methodology is the following: after somebody enters the apartment, there is motion to be expected within the first minutes while the person or people entering the house settle down. Therefore, if the next few timestamps record any movement, the occupancy state will be activated and left unchanged until the next door-opening is recorded. The number of timestamps after door-opening to consider, which corresponds to the timeout of the virtual sensor, is a parameter that needs to be calibrated. The specific value to be used depends on the parameters of physical sensors used on the site. An example of a relevant sensor parameter is motion sensor hold (i.e. how long after the last detected motion will the sensor keep the motion flag set). Different statistical insights can be built to decide it wisely.

Fig. **7** 7 shows the proposed state change diagram for the virtual real-time sensor.

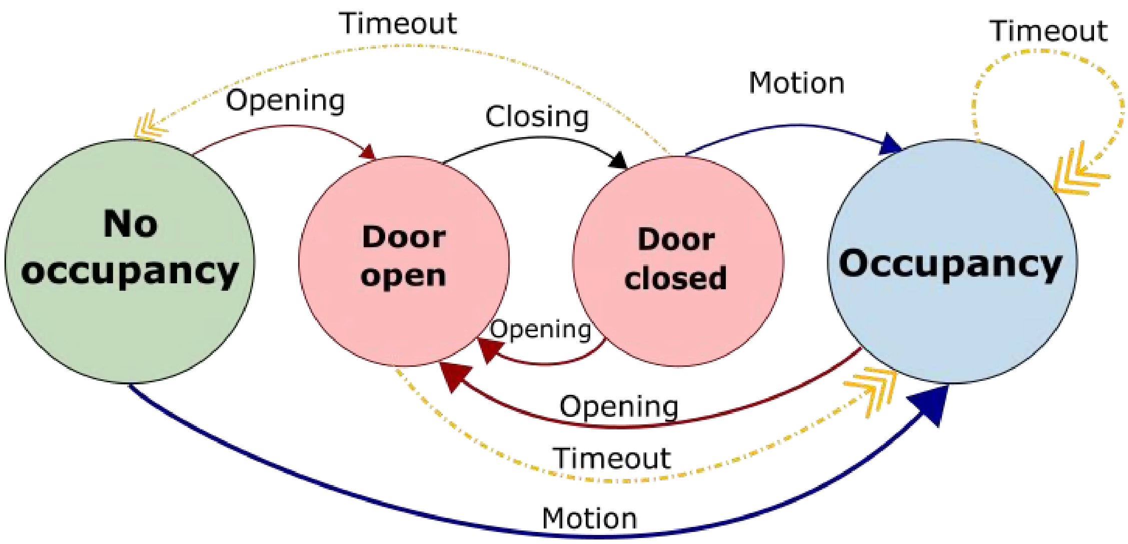

Fig. 7: State change diagram for occupancy inference

3.3. Validation and analysis

For validating our virtual occupancy sensor, as real occupancy information was not available as part of the dataset, visual inspection and statistical insights have been used for verification. First, a random apartment and a 2-day interval have been chosen to see how the aggregated motion of an apartment – aggregated by taking the mean value of all the sensors in the apartment into a single variable – door openings and inferred occupancy state look like when represented together. This can be seen in Fig. 8.

As is noticeable, the virtual occupancy sensor works as expected, the state does not change until the door openings happen and it is activated even when motion is not recorded. On the other side, even if the data used for the generation of this plot is resampled in 15-minute interval, i.e. coarser time resolution, at the timestamp after the door opening there is always motion recordings. This also shows how the online occupancy inference technique could be useful in the case of a coarser historic dataset.

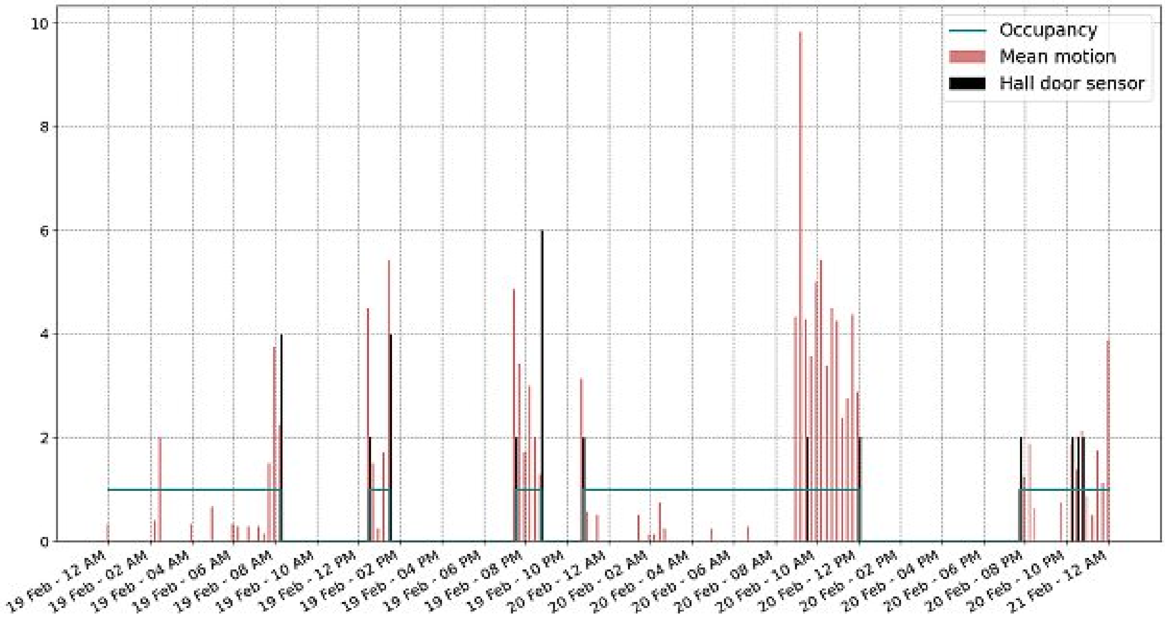

Fig. 8: Visual representation of the inferred occupancy info from mean motion and entrance door opening data

For getting a better insight how the online occupancy inference would work, it has been simulated with the current dataset. We have done this by calculating the ratio between the number of door openings which are followed by a non-zero motion in the next two timestamps and the total number of occupied states which follow a door opening. Therefore, a value of 1 would mean that the online methodology works as well as the historic data processing method and 0 would mean the opposite. Tab. 1 shows for 10 randomly selected apartments how these ratios are in general very close to 1, which means that online methodology is almost as accurate as the historic data processing method even when the resampling interval is very high for such an assumption to work.

Tab. 1: Ratios calculated dividing the number of door openings which are followed by motion in the next timestamp by the total door openings which are followed by occupied state

Apartment	**636**	**421**	**214**	**783**	**318**	**117**	**633**	**136**	**255**	**250**
Ratio	0.93	0.90	0.98	0.86	0.97	0.98	0.97	0.99	0.92	0.97

3.4. Possible drawbacks

Even if the methodology has worked for the use case we worked on, there are some general problems that could interfere with the inference. First, in our case no threshold has been imposed to consider a group of motion recordings as movements being made by a human occupant. Therefore, if there is a pet in the household, the motion sensor is placed in a way that could detect its movement and the occupancy detection does not want to consider pets as occupants, the result would be incorrect. Some motion sensors have an option to calibrate the sensitivity, so that they do not detect motion of smaller objects.

Other common problems include hardware issues, where e.g. door recordings cannot be sent to the database due to a failure of the sensor or a loss of connectivity. In this case, both methodologies would have similar problems assigning correct occupancy states in time intervals when hardware issues occurred. As previously mentioned, several methods for identifying data gaps have been developed that can detect such issues. One of the ways of dealing with them is to assign an "undecided" state for those periods, so that possible automations that depend on correctly detected occupancy information are not activated erroneously.

4. Relevance for Facility Management

Occupancy information can provide facility managers insights about the people flow dynamics and recurring patterns within a building. The methodology can be extrapolated to all kinds of buildings for which there is populated semantic knowledge base as well as all the necessary sensors.

For office buildings, the historical data of room occupancy can be compared with the expected or reserved room capacity, for example by analysing if room bookings are actually used or if there are vacancies in the reserved rooms, allowing for a reduction of rented office space. Depending on the location of the motion sensors, motionless workers cannot be detected using just motion sensors - our methodology proposes a solution to this shortcoming of commonly used motion sensors. In the case of hotels, this privacy-preserving and cost-effective occupancy inference methodology could be useful for monitoring rooms to know whether to dispatch cleaning staff without the need for customers to put "do not disturb" signs. It could also be used in automation scenarios to safely turn off heating or cooling equipment turned on by the occupant that has left the room. Finally, in residential buildings several safety-related scenarios are imaginable where non-occupancy in connection with activated high-energy devices such as heating bodies or stoves and ovens could be cross checked to detect dangerous situations and cut power or alarm the designated persons.

5. Conclusion and future work

Further integration of available sensor information with topology semantic knowledge will continue to improve the operation quality for technical facility management. Research continues in the direction of sensor fusion, by putting together virtual sensors that use different data source like temperature, CO_2, door contacts, illuminance and other existing sensors. Merging this information can improve indoor air quality, comfort, energy efficiency, and especially operation costs. Occupancy state is relevant for different industries in the building and should therefore be a central data source for HVAC systems, room management and security applications. We see a potential for usage of detected occupancy in several areas. Occupancy state can be used as a more accurate input for building more accurate machine learning occupancy models. More accurate occupancy models can be further leveraged for HVAC system operation optimization scenarios, such as preheating. Furthermore, detected occupancy state could be directly used for automatic turn off on no occupancy. Our simple method, connecting semantic knowledge with sensor data, provides a robust occupancy state detection which can be used in historic or streaming data processing scenarios. When required time series are already available, replication of the method is cheap and simple, and required manual work is limited to describing the related spaces, locations of sensors and links to collected time series using a predefined ontology. As future work, the idea is to use this feature to improve energy consumption prediction models using different aggregated and processed forms of it, as cumulative presence or via a pattern analysis.

ACKNOWLEDGMENT: This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 768776, project Inbetween (ICT enabled BEhavioral change ToWards Energy EfficieNt lifestyles).

References

Beetz, J., van Leeuwen, J., & de Vries, B. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1), 89–101. https://doi.org/10.1017/S0890060409000122

Lin, X. & Lau, J. (2014): Demand controlled ventilation for multiple zone HVAC systems: CO2-based dynamic reset (RP 1547). HVAC&R Res. 20 (8) 875–888.

Mir, M.U. (2018): Room occupancy estimation for the building automation, Master Thesis at FH Joanneum, Univ. of Applied Sciences

Rasmussen, M. H., Pauwels, P., Hviid, C. A., & Karlshøj, J. (2017). Proposing a central AEC ontology that allows for domain specific extensions. Proceedings of the Joint Conference on Computing in Construction (JC3), 1, 237–244. https://doi.org/10.24928/JC3-

RDF - Semantic Web Standards. (n.d.). Retrieved July 10, 2020, from https://www.w3.org/RDF/

Šipetić, M., Jentsch, R., Aizpuru, J., & Kurzidim, J. (2020). Design and integration of the project-specific ontology for data analytics support. Proceedings of the 8th Linked Data in Architecture and Construction Workshop (pp. 50-63). Dublin, Ireland: CEUR Workshop proceedings.

Ürge-Vorsatz, D., N. Eyre, P. Graham, D. Harvey, E. Hertwich, Y. Jiang, C. Kornevall, M. Majumdar, J. E. McMahon, S. Mirasgedis, S. Murakami and A. Novikova (2012). Chapter 10 - Energy End-Use: Building. In Global Energy Assessment - Toward a Sustainable Future, Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria, pp. 649-760.

Yang, Z.; Li, N.; Becerik-Gerber, B.; Orosz, M. (2014). A systematic approach to occupancy modeling in ambient sensor-rich buildings. Simulation, 90, 960–977.

Zucker, G., Sporr, A., Garrido-Marijuan, A., Ferhatbegovic, T., & Hofmann, R. (2017): A ventilation system controller based on pressure-drop and CO2 models. *Energy and Buildings,* vol. 155, no. Supplement C, pp. 378-389, doi: 10.1016/j.enbuild.2017.09.041