Network as a sensor for smart crowd analysis and service improvement

Dr Mu Mu, University of Northampton, Northampton, NN1 5PH, United Kingdom

Abstract

With the growing availability of data processing and machine learning infrastructures, crowd analysis is becoming an important tool to tackle economic, social, and environmental challenges in smart communities. The heterogeneous crowd movement data captured by IoT solutions can inform policy-making and quick responses to community events or incidents. However, conventional crowd-monitoring techniques using video cameras and facial recognition are intrusive to everyday life. This article introduces a novel non-intrusive crowd monitoring solution which uses 1,500+ software-defined networks (SDN) assisted WiFi access points as 24/7 sensors to monitor and analyze crowd information. Prototypes and crowd behavior models have been developed using over 900 million WiFi records captured on a university campus. We use a range of data visualization and time-series data analysis tools to uncover complex and dynamic patterns in large-scale crowd data. The results can greatly benefit organizations and individuals in smart communities for data-driven service improvement.

Introduction

ICT infrastructures have a critical role in smart community initiatives where organizations, residents and visitors use digital information for positive impact. Many services in our communities are planned in advance but never optimized based on real-world usage or on-demand requests. Crowd monitoring can provide the information needed to enhance service intelligence. Organizations can capture and evaluate how visitors access different parts of their premises and how such data evolve over time. The information can be used in data-driven risk assessment and resource planning for improved quality of life. Crowd monitoring also provides critical information for smart energy solutions to reduce energy consumption and carbon emissions in buildings and communal areas. Without ICT infrastructures, crowd monitoring is normally carried out by human operators which is inaccurate and inefficient. Many modern crowd-monitoring solutions have been designed to use video cameras paired with computer vision tools or LiDAR to monitor social distancing and crowd movements without human intervention [1] [2]. Unmanned aerial vehicle (UAV) has also been used for abnormal crowd behavior detection with the help of edge computing and machine learning [3]. Other mobile-based solutions such as contact-tracing apps use Bluetooth events to estimate the proximity between mobile phone carriers [4]. A survey conducted by Singh et al. compares a range of technologies used for crowd monitoring [5] and it shows that the majority of the solutions in operations are still based on video cameras due to their superior accuracy.

Related work in the past has shown various attempts to use WiFi signals for crowd monitoring in different application scenarios. Chilipirea et al. installed 27 outdoor WiFi scanners on city streets and used WiFi probing signals from mobile devices to analyze crowd data during a public event [6]. In this scenario, mobile devices are not connected to any WiFi network. The authors discussed the challenges of extracting crowd information in a dynamic environment using sparsely distributed scanners. Determe et al. also investigated the use of WiFi probing packets to estimate crowd information in the indoor environment and compare their results with visitor counts from a video camera feed [7]. In [8] the use of WiFi probing for people counting and mobility detection was expanded to edge networks using commercial WiFi scanners. The solution shows its practicality despite the lower precision compared with camera-based solutions. Braham studied a WiFi dataset from 10 access points placed in the centre of City Enschede [9]. The author explored how different movement patterns such as the ways that crowds approached the city centre from different surrounding areas, can be inferred from the WiFi connectivity data. Due to its limited number of access points, the dataset cannot offer the location accuracy needed for smart applications. Overall, existing crowd-monitoring solutions are highly dependent on specialised equipment such as surveillance cameras and high-resolution sensors which can be costly and intrusive to everyday life. Current WiFi-based solutions are advantageous for practicality and privacy but lack the accuracy to support smart communities.

This article introduces a WiFi-based crowd-monitoring solution which uses network telemetry over 1,500 WiFi access points (APs) to form a network of always-on sensors. It takes a fundamentally different approach compared with previous WiFi probing-based solutions. Our work offers the following advantages compared with existing solutions:

- It only uses data from standard WiFi services and does not require WiFi scanners.
- It can track the movements of individual user devices for high-precision indoor and outdoor crowd monitoring and predictive analysis.
- It can associate multiple devices that belong to the same user to further improve the accuracy of people counting.
- Our solution includes tailored interactive web applications to visualize valuable crowd information for community users.
- Our work is built upon a live dataset of over 900 million WiFi records from a smart campus network.

The remainder of the paper introduces the supporting systems and a range of applications developed for smart community solutions.

System Overview

Network infrastructure for monitoring

Figure 1 illustrates the WiFi-based crowd-monitoring system deployed at the University of Northampton's Waterside campus. The campus has a state-of-the-art software-defined network (SDN) infrastructure to provide internet connectivity to its residents and visitors. SDN champions the new network paradigm that separates network control from the data forwarding functions. This allows network equipment from different vendors to be managed by a single network controller as long as they conform to the same SDN specifications. SDN has been increasingly adopted for the Internet of Things (IoT) and intelligent service applications [10]. The campus uses Cisco DNA Center as its primary network controller for network management and data aggregation. With its network telemetry features, the controller uses a set of automated processes to capture information about the connected devices. Around 1,500 indoor and outdoor APs provide wireless internet access to mobile user devices across the campus. Each AP connects to a network switch via a high-speed wired connection and provides Wi-Fi connectivity in a designated area. Extended WiFi coverage is achieved by strategically placing APs. User devices moving across physical locations will automatically detect available APs in the area and switch between APs seamlessly based on signal strength.

Most buildings at Waterside campus have between 20 to 55 APs installed on each floor depending on the functions and the internal structures of the building. Most rooms have at least one dedicated AP while some large meeting rooms, foyers and corridors are equipped with several APs to warrant excellent WiFi coverage. There are over 30 outdoor APs for the outdoor workspace and for continuous internet connectivity while visitors travel between buildings. The physical location and installation mode (e.g., ceiling, wall, post) of each AP are registered for maintenance purposes. For instance, one AP is marked on the floor map with the descriptive text “outside room 210” and its installation type is “ceiling”. Using SDN, a single network controller manages all network switches and wireless APs. This enables a live and unified view of all devices connected to each AP. The density and the distribution of APs in the physical space determine the accuracy of crowd monitoring. The topology of the access network such as switches and routers does not impact the effectiveness of crowd monitoring because their role is to relay AP information to the network controller for data aggregation. In summary, our solution uses information readily available at network controllers and does not require changes to the network.

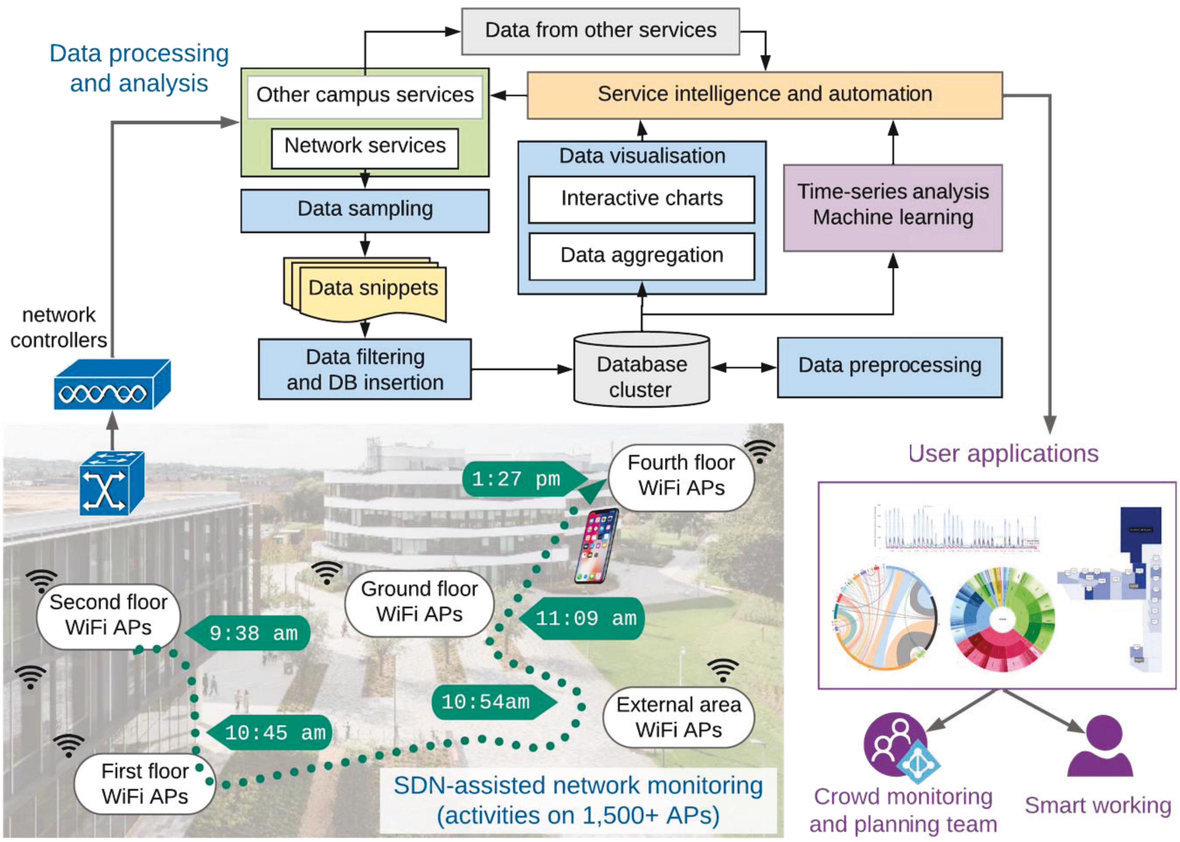


Figure 1 smart campus system

Data capturing and processing

In order to study the crowd dynamics over time and to support predictive analysis, live device information at each AP must be sampled and stored. Each sample is a snapshot of all wirelessly connected devices and their associated APs. The number of connected devices is usually in the range of 1,000 to 6,000 depending on the time of the day and the day of the week. A dedicated server takes one sample every 60 seconds. This sampling rate was chosen based on four factors: how frequently devices switch between APs, the balance between noise and fidelity, the impact on the network controller APIs, and the storage capacity. In practice, visitors are either stationary or moving at regular walking speed. A sampling rate too high will lead to a significant amount of redundant data and a sampling rate too low will not capture details of the movements e.g., whether a change of location to a different floor was made via staircases or an elevator.

Overall, the data collection process gathers over one million records per day. The raw data include over 20 data fields including *datetime*, *MAC address*, *IP address*, *WiFi SSID*, *user ID*, *device type*, and *WiFi protocol*. A data processing function goes through all unprocessed samples and populates a database asynchronously with filtered and formatted data. Visitors may have more than one device connected with the same user ID. By associating time-coded records of connected devices and the geospatial information of APs, the system can discover how crowds are distributed on-site and how their locations change over time. So far over 900 million records of data have been collected since the beginning of our smart campus project. Table 1 lists sample records based on the raw data captured from network controllers.

Table 1 Illustration of data captured from network controllers

date time	MAC address	IP address	AP name	network name	user ID
02/07/2020 10:00:01	b6773f96309c5aa22a81ee034ff7f559	a77803ead788503528d8c3b7f4650798	AP-MB-073aca	UoN_Guest	b475a86b6789657cb91aec5d35199e1f
02/07/2020 10:00:01	c6a54e36752ad3fb8bdd8789fc48f78a	eddc6c827a4efd013f3eb2d8c83da70f	AP-PA-ce6243	UoN_Student	b07d4a81b42a55843ad90372891afdd4
02/07/2020 10:00:01	a3bd6f7c930955aa8041006a87822c6c	2e5df82147720bd9a6c06341085e5163	AP-PA-a7a404	UoN_Student	b766002b992e6b9126a860c121322764
02/07/2020 10:00:01	a1f4da2723d3abeeb4ffdea5725fe9fe	9dc8161de5d5db9f2128d44cf7a1ca09	AP-FC-be6d34	eduroam	9fd88b12be4e29077c812f67f1298563
02/07/2020 10:00:01	f3952fc8b937e46ccb2c546809000c7a	3e5e5cf2a7d91a60c59a52694cdbafaf	AP-SP-e3d04e	SSN	0ca55239ec92e00a8f241827aab87197
02/07/2020 10:00:01	a24a7eb47cb3fc0f2cf1894b471c61ad	ba1547ba8a3d2c137f53a3437bd97d00	AP-SP-e3d04e	SSN	0ca55239ec92e00a8f241827aab87197
02/07/2020 10:00:01	e6b3f04847a3d6ec6ca0838efb17344c	e3fe301f433c36584f4144ee71619e16	AP-LH-b71103	UoN_Guest	1bab53c8e7701ae19cdf9c0600b958b0
02/07/2020 10:00:01	553fe2b11b0276ff3e56606775bd17d5	5a72482963b47052cf199385e92b74f7	AP-EX-4c6223	UoN_Staff	7a2db8b5da9df1af718dac328aa82ab2
02/07/2020 10:00:01	bf340ebab92ffd058fbca491fe8a810e	1498188b8e522b3f1b2f3aced830b29c	AP-LH-b71103	UoN_Staff	afd57ae92e3bb268a44bcba932f501d0
02/07/2020 10:00:01	92d281f900caea374c36840222d6c403	43cb1fb6336fa16bf1f670db4f6a2610	AP-PA-ca35c8	UoN_Student	3eb93fd0c0fc0e3af148cec22b64de0c
02/07/2020 10:00:01	3438b08e88caa00e74ff0b2cd49573af	039b663892529e2e9f63866df08d24e3	AP-FC-c55812	UoN_Gaming	3f6e9a7cb490f8451ed4e25644d6edde
02/07/2020 10:00:01	2d58639d0ebf3c6b2f65ac60625bf1af	eea5d4f7d98fa3444323fc8942a5ef53	AP-EX-e74152	UoN_Guest	c1221a0cc64bb3cc3afc83563ff40cc4
02/07/2020 10:00:01	8f2629130314115ab3e3f19fc232fd07	ca0a2f88b18695fe964b10f8da7f47bb	AP-MB-2f9022	UoN_Student	2fc002cecbd93ed4a25bbe2cf9a34887
02/07/2020 10:00:01	c867421c3c34d10214f68e70cf4eb37c	a2b7c90f6021c5cdc6458853ecee7f7d	AP-PA-692f05	eduroam	e15828712cecc38830ee99c6a35cc1be
02/07/2020 10:00:01	b9f66a7c0afb2e365d371e0f0c02045a	3556a889d95f627bda5bf529fb0d87b6	AP-LG-e381d8	UoN_Staff	e917a1d584bc134b96923bb59cdaa167
02/07/2020 10:00:01	16bbe07b42320402b688545c248c7f29	e06227e4476ea0b2ff1203c0102d0621	AP-PA-227f03	UoN_Guest	0ca55239ec92e00a8f241827aab87197
02/07/2020 10:00:01	c1339963cc359bda370cbc09adbec732	bbda7c686b458b756ff8d268ff39129b	AP-PA-fed0cd	UoN_Guest	f7939a5171db387f7c525ef7944bb4f2
02/07/2020 10:00:01	e0dfc56530be09297b24a5cb4d6ddf11	8d02661662fcd564105a2189b78a9711	AP-PA-5ed3b6	UoN_Student	eb9b08c7d9e9c3fd3ab59204e75575a8
02/07/2020 10:00:01	75e15c13d32bac8493184ee25e597dd1	1f1635aecbb75b6186365b9e35072e25	AP-CB-bb81e8	UoN_Student	1dc108de42b6f0c11a4853db319e8c59

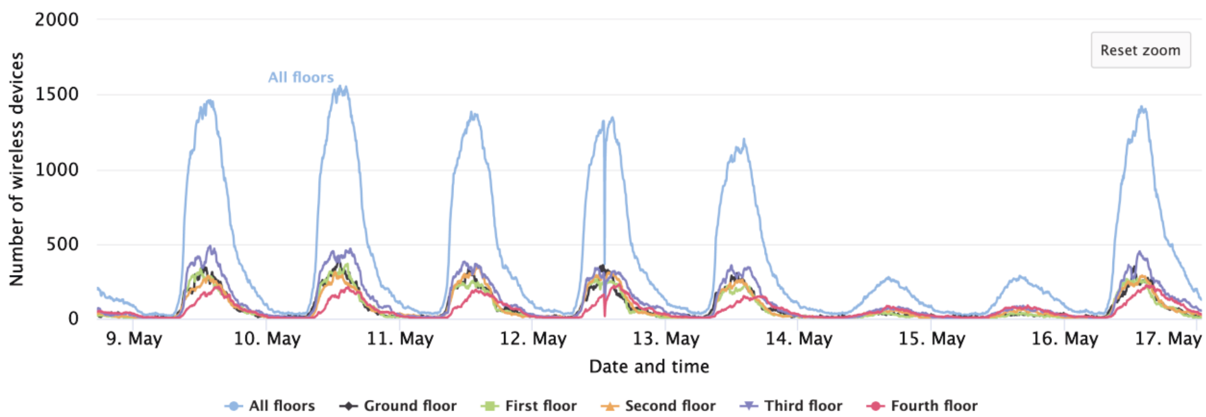
Personal and device information is anonymized via a custom hashing function as part of the sampling process. Due to the low probability of hash collision, the hashed data are considered unique and distinguishable. This allows us to estimate the number of visitors (unique user IDs) in an area (defined by the coverage of one or multiple APs) and how each person moves between areas.

A range of data analysis and visualization applications have been developed to extract high-level information from the large volume of collected data. The aim of the applications is to inform decision-making at both organisational and visitor levels. The original use cases of the project were promoting social interactions and closeness, energy efficiency, safety, and timetabling. To support COVID-19 recovery, many features have been extended to support safeguarding and social distancing. Examples of COVID-19 use cases are crowd density risk assessment, proactive workspace planning and scheduling, event simulation, and lone worker support. The following sections discuss user behaviour modelling and the developments made to support the aforementioned use cases.

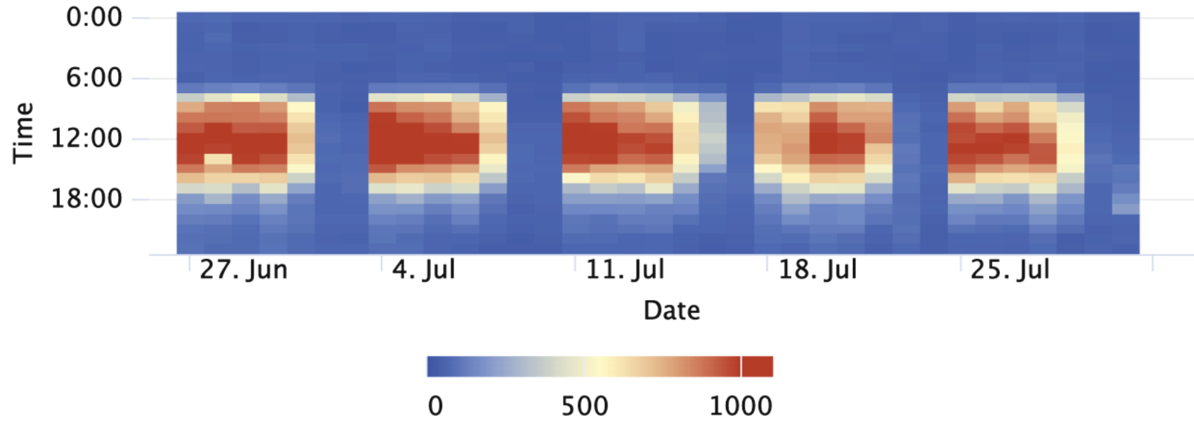
Crowd dynamics

Figure 2 shows the crowd dynamics over time in multiple communal areas. These figures were captured from a web portal that was designed for the services and facilities team to access crowd information. Users can use interactive features to zoom in on a selected time period or select any data label for further examination. Figure 2 (a) gives the crowd activities inside a building and the breakdown data across its five floors. Knowing how visitors access different levels of a building provides valuable insights into how resources should be distributed to accommodate service requirements. It also provides a quantitative measurement to assist fire and safety assessment at a typical time and day of the week. With longitudinal crowd data, the security team could also detect anomalies in access patterns and unusual behaviours such as substantial activities outside working hours.

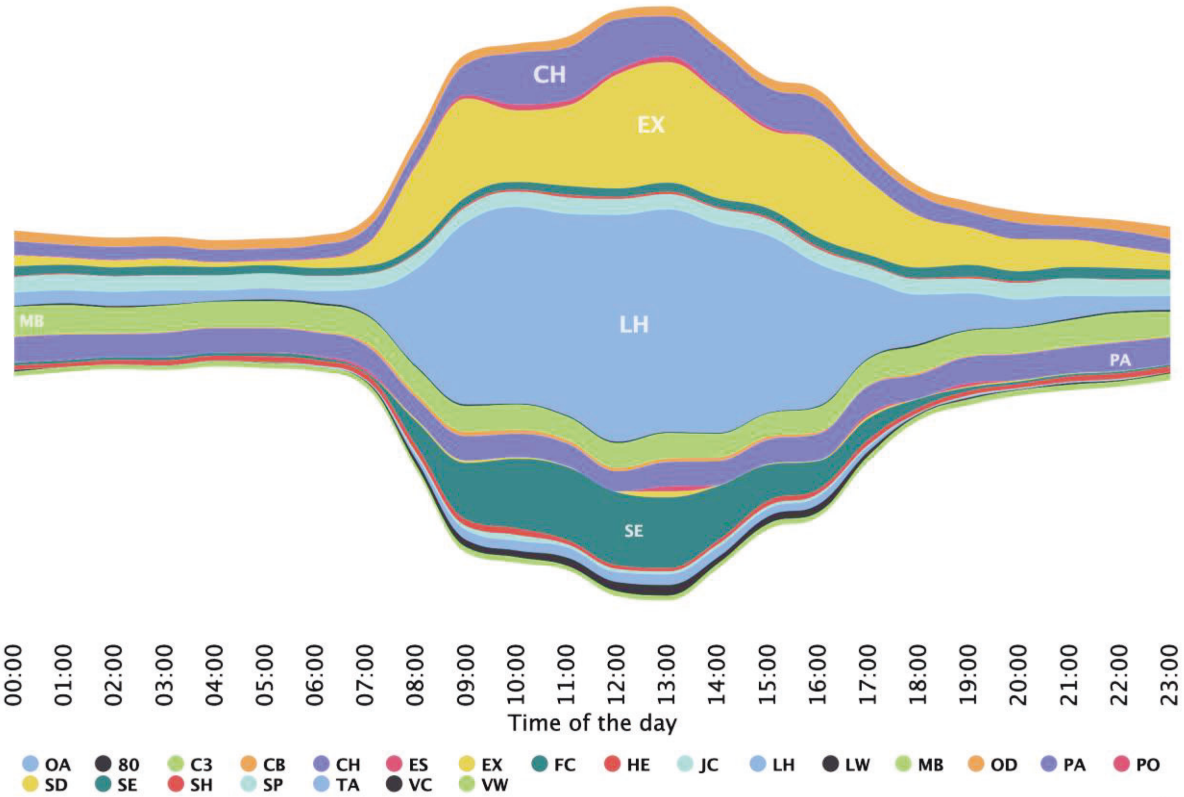
Figure 2 (b) uses heatmaps to illustrate the dynamics in population density over five weeks in a given area. The heatmap is quick and easy to understand without the need to read detailed figures. It helps service teams optimize resource allocations based on historical and live data. For instance, Mondays and Fridays show significantly different service demands that need to be catered to differently. The campus sees very few visitors during weekends but on Saturday 16th of July, a social event attracted hundreds of visitors. For a large public space with multiple communal areas, it is important to observe and compare crowd density at different granularity levels at the same time. Figure 2 (c) is a stream graph that shows how the population density changes on a single day and how the population was distributed across multiple areas. Crowd density from different areas is colour-coded and stacked around the central axis. For instance, the light blue area in the center depicts the number of visitors in LH area, a large multi-functional 5-story building. From 7:00 am, the crowd started to arrive on site, as can be seen from the growing number of wireless devices. Most crowds went to the four most prominent areas where the majority of the services and office space are allocated. Most areas saw their population peak in mid-day. The crowd in LH area stayed later compared with other areas such as SE where most visitors seem to depart before 18:00. This information is useful to enable smart planning of security and support staff rota in late afternoons and early evenings when resources are limited.



(a) crowd activities across five floors within a building



(b) crowd density heatmap



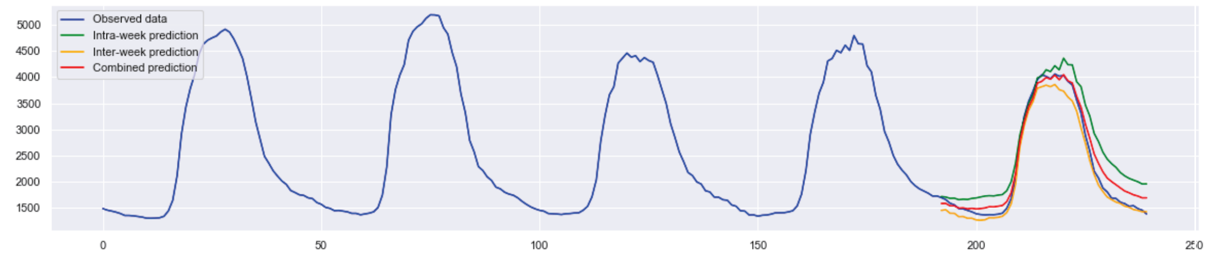
(c) stream graph of crowd activities across multiple areas

Figure 2 Crowd dynamics

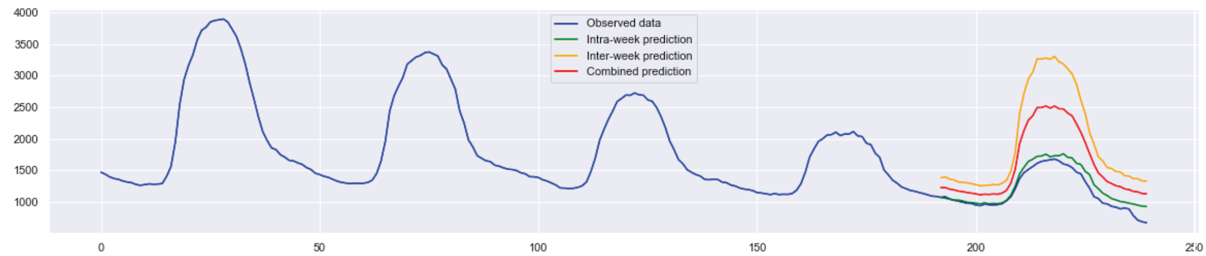
Crowd modelling and predictive analysis

The capability to predict the crowd density at a particular point in time is an essential tool for proactive planning and anomaly detection in smart communities. This will help build a more efficient and safer environment for community members. Our crowd information is time series data as they are sequences of numerical data points in successive order observed at regular intervals (i.e., every 60 seconds). Time-series analysis enables us to model trends and patterns which can support the prediction of the value at a future point. ARIMA (Auto Regressive Integrated Moving Average) [11] is a common time-series analysis method characterized by 3 terms: *p*, *d*, *q* which capture the pattern of changes in the data ("auto-regressive"), the rate of changes in the data ("integrated") and the noise between consecutive time points ("moving average"). The seasonal variation of ARIMA (SARIMA) introduces additional terms to capture seasonal differences for non-stationary data [12]. Our data are non-stationary with multiple levels of seasonality embedded. Taking a top-down view: the institution organizes most of its activities in terms, each term is comprised of weeks with special events at the beginning and the end, each week has weekdays with many visitors and weekends that see few visitors, and each weekday includes regular working hours. Using different season configurations, the SARIMA can capture seasonality patterns at different levels.

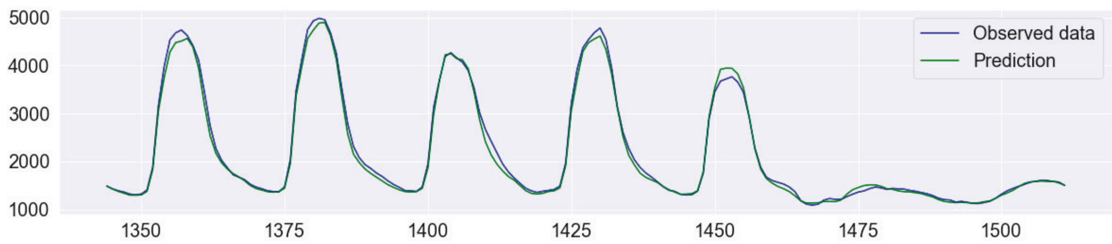
We first used the SARIMA method to predict crowd level on a single day. Hence, the seasonal term was configured to cover observations from a day. Two types of predictions were tested: intra-week prediction and inter-week prediction. The intra-week prediction uses data observed on the previous days of a week to estimate data on future days in that week, e.g., using data from Monday to Thursday to estimate the data on Friday. The inter-week prediction is based on the data associated with the same weekday but from previous weeks, e.g., using Fridays from several weeks in the past to predict data on future Fridays.



(a) crowd dynamics during a typical week



(b) crowd dynamics when people were encouraged to work from home



(c) week-level prediction of crowd dynamics

Figure 3 SARIMA predictions

Figure 3(a) shows the predictions of the crowd level on Friday, March 6 of a normal term week. The blue curve visualizes the observed data for that week (Monday to Friday). The green curve shows the results of an intra-week prediction where data from four days of the week was modelled. The intra-week prediction slightly overestimates the crowd level for

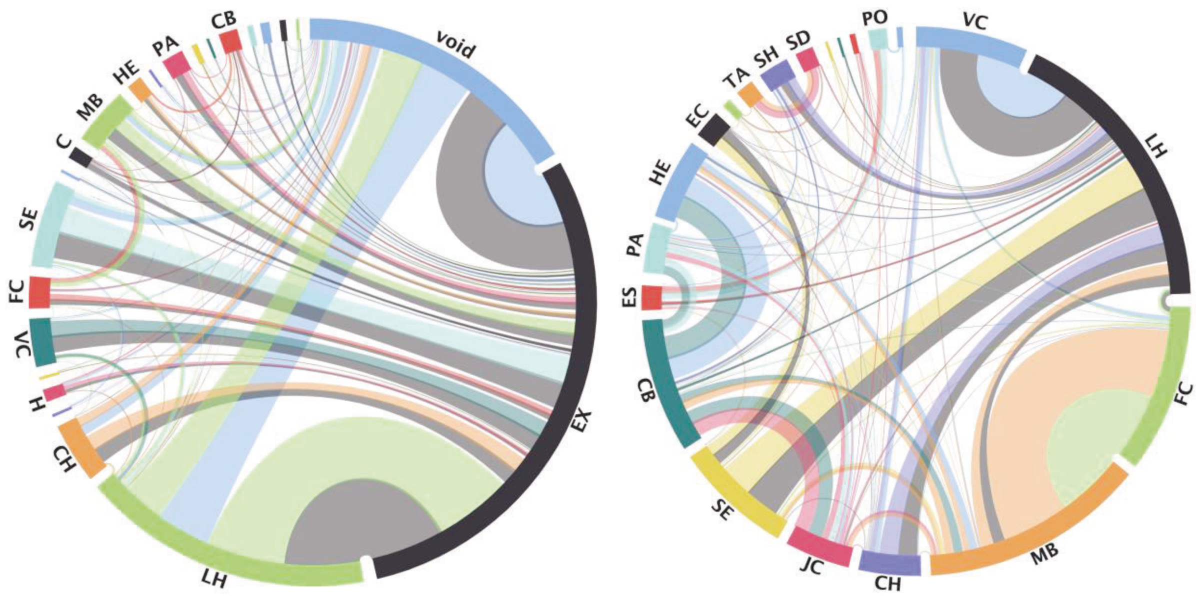
Friday afternoon. We believe this is attributed to people leaving early on Friday afternoons compared with other weekdays (as observed in Figure 2 (b)). The intra-week prediction cannot capture that extraordinary weekly pattern. The orange curve gives the inter-week prediction results based on the previous four Fridays. This method captures normal activities on Fridays but is agnostic to week-specific changes (e.g., public events or facility closure for maintenance). On this particular Friday, staff and student ambassadors were doing additional work preparing for a public “Discovery Day” for prospective students and their families. This explains the slight underestimation made by the inter-week prediction. Balancing intra- and inter-week predictions using a simple element-wise averaging function, the red curve shows the results when the outcomes from the two predictions are combined with equal weights.

To investigate how the SARIMA-based models capture the impact of the COVID-19 pandemic on crowd density, the second case study chose the week prior to the COVID-19 lockdown (Figure 3(b)). This week is considered an “abnormal” week as crowd behaviors deviate from normal work patterns whilst people spent more time studying or working from home. In this case, the intra-week model successfully captures the decreasing trend of visitor numbers in that week, when other methods generally over-estimate the crowd based on previous weeks.

By adjusting the seasonal configuration, the SARIMA method can model and predict data in the unit of a week. Using data from 7 consecutive term weeks, a model is constructed to predict the crowd level of the following week. Training data were down-sampled at 30-minute intervals to speed up the modelling process. Figure 3(c) depicts the results of the week-level prediction. The model accurately captures the unique crowd characteristics on different days of the week.

Cross-area movements

Crowd movements between areas are often determined by how physical facilities were designed and how events such as meetings and classes are planned. In some cases, cross-area movements are encouraged for the benefit of the physical and mental health of community members or to facilitate innovation across boundaries in large organizations. There are also scenarios where extensive travelling between areas is considered a sign of poor planning and scheduling. For instance, simulations of crowd movements are often carried out to evaluate whether a public venue has sufficient access routes to allow a large number of visitors (flash crowd) to safely arrive or leave simultaneously. While combating COVID-19, unnecessary movements between floors or buildings may be discouraged as there are often “bottlenecks” such as stairs, elevators, and building entrances when people move between areas. Social distancing is less likely to be kept at the “bottlenecks”. Therefore, monitoring and understanding cross-area movements can assist risk assessment at high-risk locations.



(a) movements between all areas (b) movements between buildings only

Figure 4 A dependency graph showing crowd movements

We developed data analysis functions to examine cross-area location changes taken by visitors. Figure 4 (a) shows a dependency graph of how visitors move between areas and buildings. The shortcodes at the edge of the graph represent different areas. For instance, “LH” stands for Learning Hub, a building whose data were used in Figure 2. “EX” is the code for all outdoor areas covered by outdoor (external) WiFi APs. “void” is a pseudo area code used to represent any out-of-campus areas. It helps visualise where devices were first observed. Normally, a mobile device with its WiFi turned on would first connect to outdoor APs (“EX”) as visitors approach the campus from car parks or footbridges. This is illustrated by the light blue outward link from “void” to “EX”. “EX” bridged most cross-building movements as visitors travelled between different parts of the campus. Figure 4 (b) shows the movements between buildings only. It gives a clear indication of how visitors access facilities distributed between multiple buildings. It can be used to assess the capacity of footpaths against real usage.

Smart floor maps

For work planning and scheduling, it is essential to monitor and evaluate how facilities such as meeting rooms, work areas and catering areas are used. Our aim is to create a floor heatmap that visualizes crowd density from live, historic, or simulated data as shown in Figure 5. In practice, each area may be covered by one or multiple APs. Based on geospatial information of APs, area-to-AP mapping is defined to concatenate network telemetry data associated with multiple APs within an area to estimate the area crowd density. For instance, the number of visitors in a large meeting room equipped with two APs can be measured based on the number of unique devices (hashed MAC addresses) or unique users (hashed user IDs) connected to the corresponding two APs. Mappings are stored in a relational database as records of *Theme ID*, *WiFi AP ID*, *area code*, and *area type*. Such mappings are intent-based and not exclusive. Different mappings can be constructed for different analysis purposes in a smart community and an AP can be associated with different mapping schemes.

For crowd density analysis, it is important to differentiate between stationary crowds (people staying in an area and purposely using the space) and crowd traffic (people moving through one area to access a different area). A person may pass by multiple areas before reaching her destination. This can leave a trace of a “digital footprint” when her user device joins and leaves multiple APs on her path. In an institution with regular timetabled events, crowd traffic can generate significant “noises” in the crowd data. For instance, a group of 50 people leaving an indoor event and exiting the building would cause a major fluctuation in crowd density measurement in all areas between the room and the exits of the building. For our use case, the analysis is centered around stationary crowds. Hence, we filtered out presence data that were observed for less than one minute by any AP. This was done by comparing every two consecutive samples and keeping only the data records that appeared in both samples.

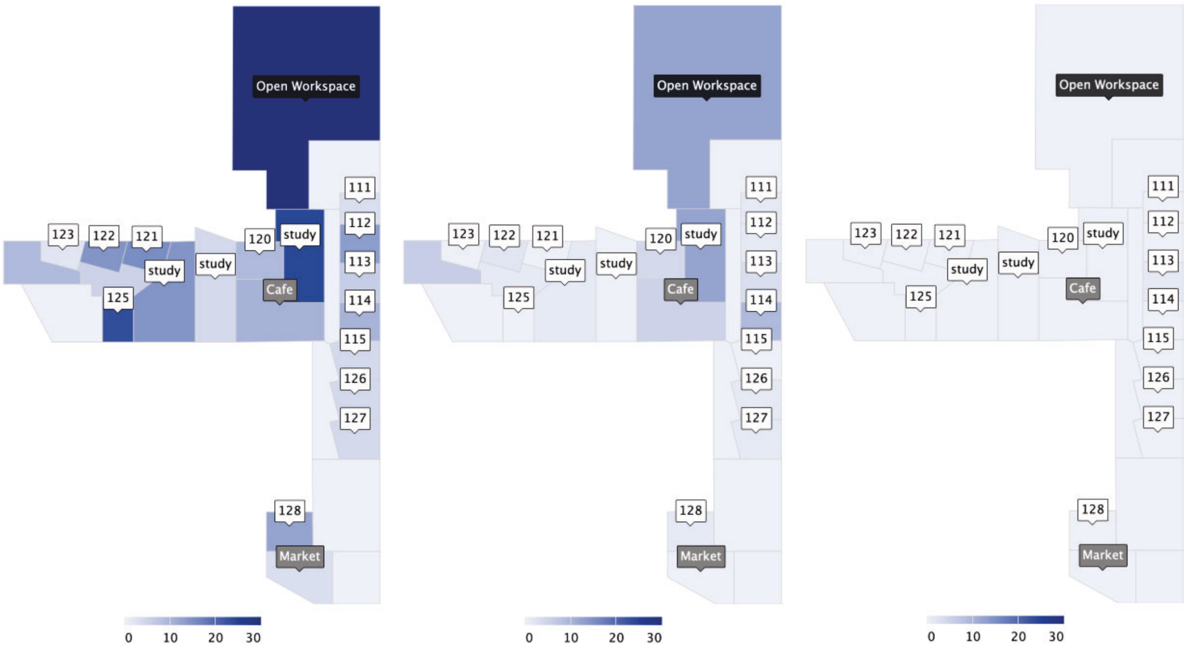


Figure 5 A comparison of area crowd density of a building floor at 10:30 am on three different days. From left to right: Pre-lockdown baseline, Near-lockdown reduced social contact, and working/studying from home during lockdown

Figure 5 shows the crowd density floor map of one building floor at three different points in time. All diagrams were developed based on the actual floor plan provided by the architect. This floor has eighteen lecture rooms and study areas (white label), one large open-plan workspace (black label), and two catering areas (grey label). A deeper shade of blue color indicates a higher number of occupants in the area. The data is based on a 10-minute average of the live data observed. This augmented floor map is part of a web application that automatically refreshes to show up-to-date crowd information. It is also able to "replay" historic data at a configured playback speed.

The comparison between the three diagrams in Figure 5 provides a unique view of the impact of the COVID-19 lockdown. The one on the left shows the data observed on Thursday, March 12, 2020 at 10:29 which is part of a week not significantly affected by COVID-19. Visitors were still using meeting rooms and study areas actively. The floor map in the middle shows the data captured one week later at the same time when many people start to work from home voluntarily. Some people still chose to access the shared space, but the use of confined spaces especially small meeting rooms had lessened significantly. Open areas became popular among visitors. The last floor map shows a deserted building floor during the lockdown period. Only a small number of authorized staff had access to the building for maintenance and security check.

Potential Applications

The paper gives the details of a data capturing and processing pipeline that allowed us to correlate time-coded device information from APs across campus and how live information of crowd density and movements can be derived at different granularity levels for a range of smart services. Multiple data visualisation methods are used to provide human-friendly graphics to assist decision-making. Predictive models are also developed to enable simulation and anomaly detection.

Besides the prototyped solutions, we envisage a range of potential applications for smart communities. The time-series modelling is particularly useful for anomaly detection at either the global or local level [13]. Automated anomaly detections allow smart communities to advance their intelligent use of resources for improved service performance and resource efficiency. The inter-week model shown in Figure 3(b) can detect global anomalies when the observed pattern significantly deviates from SARIMA's inter-week prediction. The service team can configure an anomaly threshold to trigger immediate responses. The local-level anomaly could be observed when there is a spike or a sudden drop in crowd population within a short period of time. This is normally measured using tools such as Exponentially Weighted Moving Average (EWMA). Causes of local anomaly include emergency evacuation, unusual access outside of standard business hours, and social events. One fire evacuation drill could be observed in Figure 2 (a) on 12th May when visitors left all floor areas in a very short period of time and returned soon after.

The cross-building movements shown in Figure 4 can be modelled using data science tools such as a Markov model. A potential use case of such modelling is crowd simulation for planning and risk assessment. The crowd behavior model can help generate a large number of simulated visitors to evaluate the bottleneck of physical space design and indicates where visitors need to most support. The crowd density floor maps are intuitive to read. They can fit squarely on large-format public displays that are strategically placed across the physical space for visitors. When augmented with overlay information, smart floor maps play a key role in helping visitors gain confidence to return to work and discover less populated areas. The solution can also be used as a persuasive technology for positive changes in human behaviours. For instance, a health and well-being application may visualise the live usage of staircases and elevators as social gameplay to persuade more people to use stairs. APs can be clustered based on their proximity to staircases or elevators to showcase any increase in staircase usage. The heatmap design (Figure 2 (b)) can be extended with additional overlay information to help adapt air-conditioning, heating or lighting in different areas based on the estimated demand.

Conclusions

Crowd monitoring and analysis have a pivotal role in future smart community designs. We have witnessed how informed risk management and policymaking can mitigate the impact of COVID-19 and help contain new outbreaks in the future. Crowd-monitoring using video cameras provides accurate results but can be intrusive to daily life. Existing WiFi-based solutions using user devices' probing signals lack the precision to support smart applications. Capitalizing on a smart campus project and an experimentation environment, a new high-precision WiFi-based crowd-monitoring solution is developed to enable live monitoring and analysis using network telemetry data from live Internet services. This project is the first of its kind that uses 1,500+ high-density indoor and outdoor WiFi APs for smart applications. Compared with traditional crowd-monitoring methods, our solution does not require specialised video cameras or WiFi-probing equipment while protecting the privacy of visitors. Developing useful smart applications based on a high volume of WiFi access data is a non-trivial task. We discussed the challenges of processing such complex data for smart communities and explored multiple data visualization and modelling tools that help reveal critical information embedded in WiFi connectivity data. For future work, we will investigate the use of WiFi signal strength measurements such as Received Signal

Strength Indicator (RSSI), a method discussed in [14] for WiFi probing solutions, to further enhance the location tracking accuracy.

An associated dataset that includes over 300 million records of WiFi access data is available at: https://bit.ly/3Dmi6X1.

Acknowledgement

This work was supported by UK Research and Innovation (UKRI) under the EPSRC Grant Software Defined Cognitive Networking (EP/P033202/1).

References

[1] S. James Fong, N. Dey and J. Chaki, “AI-Enabled Technologies that Fight the Coronavirus Outbreak,” in *SpringerBriefs in Applied Sciences and Technology book series*, Singapore, Springer, 2020.

[2] D. Yang, E. Yurtsever, V. Renganathan, K. A. Redmill and Ü. Özgüner, “A Vision-based Social Distancing and Critical Density Detection System for COVID-19,” *arXiv:2007.03578 [eess.IV],* 2020.

[3] Y. Miao, J. Yang, B. Alzahrani, G. Lv, T. Alafif, A. Barnawi and M. Chen, Abnormal Behavior Learning Based on Edge Computing toward a Crowd Monitoring System, IEEE Network, 2022.

[4] Apple Inc., “Apple and Google partner on COVID-19 contact tracing technology,” Apple Inc., 2020. [Online]. Available: https://www.apple.com/uk/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.

[5] U. Singh, J.-F. Determe, F. Horlin and P. de Donc, “Crowd Monitoring: State-of-the-Art and Future Directions,” *IETE Technical Review,* vol. 38, no. 6, p. 578–594, Nov. 2021,.

[6] C. Chilipirea, A. Petre, C. Dobre and M. van S, “Presumably Simple: Monitoring Crowds Using WiFi,” *2016 17th IEEE International Conference on Mobile Data Management (MDM),* p. 220–225. doi: 10.1109/MDM.2016.42., Jun. 2016.

[7] J.-F. Determe, S. Azzagnuni, U. Singh and F. Horlin, , “Monitoring Large Crowds With WiFi: A Privacy-Preserving Approach,” *IEEE Syst J,* vol. 16, no. no. 2, p. 2148–2159, Jun. 2022.

[8] K. Gebru, M. Rapelli, R. Rusca, C. Casetti, F. Chiasserini and P. Giaccone, “Edge-based passive crowd monitoring through WiFi Beacons,,” *Computer Communications,* vol. 192, p. 163–170, Aug. 2022.

[9] K. Braham, *Smart Cities: Crowd Management Using WiFi Based Infrastructure,* Twente: University of Twente, 2018.

[10] Z. Lv and W. Xiu, “Interaction of edge-cloud computing based on SDN and NFV for next generation IoT,” *IEEE Internet of Things Journal,* vol. 7, no. 7, pp. 5706-5712, 2019.

[11] G. Box, G. Jenkins and G. Reinsel, Time Series Analysis; Forecasting and Control. 3rd Edition, New Jersey: Prentice Hall, 1994.

[12] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice, 3rd edition, Melbourne, Australia: OTexts, 2019.

[13] P. Jinka and B. Schwartz, Anomaly Detection for Monitoring: A Statistical Approach to Time Series Anomaly Detection, O'Reilly, 2015.

[14] G. Solmaz, P. Baranwal and F. Cirillo, “CountMeIn: Adaptive Crowd Estimation with Wi-Fi in Smart Cities,” in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Pisa, Italy, 2022.

Author contact: mu.mu@northampton.ac.uk
Author bio:

Mu Mu is Professor of Human-Centred Data Intelligence at the University of Northampton, UK. He has 15 years of research experience in Quality of Experience, immersive media, and programmable networks. Mu has been the principal investigator and lead researcher of research programmes funded by European Commission and UK Research Councils.