Chapter 11
Wireless Crowd Detection for Smart Overtourism Mitigation

Tomás Mestre dos Santos, Rui Neto Marinheiro, and Fernando Brito e Abreu

Abstract Overtourism occurs when the number of tourists exceeds the carrying capacity of a destination, leading to negative impacts on the environment, culture, and quality of life for residents. By monitoring overtourism, destination managers can identify areas of concern and implement measures to mitigate the negative impacts of tourism while promoting smarter tourism practices. This can help ensure that tourism benefits both visitors and residents while preserving the natural and cultural resources that make these destinations so appealing.

This chapter describes a low-cost approach to monitoring overtourism based on mobile devices' wireless activity. A flexible architecture was designed for a smart tourism toolkit to be used by small and medium-sized enterprises (SMEs) in crowding management solutions, to build better tourism services, improve efficiency and sustainability, and reduce the overwhelming feeling of pressure in critical hotspots.

The crowding sensors count the number of surrounding mobile devices, by detecting trace elements of wireless technologies, mitigating the effect of MAC address randomization. They run detection programs for several technologies, and fingerprinting analysis results are only stored locally in an anonymized database, without infringing privacy rights. After that edge computing, sensors communicate the crowding information to a cloud server, by using a variety of uplink techniques to mitigate local connectivity limitations, something that has been often disregarded in alternative approaches.

Field validation of sensors has been performed on Iscte's campus. Preliminary results show that these sensors can be deployed in multiple scenarios and provide a diversity of spatiotemporal crowding data that can scaffold tourism overcrowding management strategies.

Keywords Overtourism · Smart tourism toolkit · Crowding sensor · Edge computing · Wi-Fi detection · Fingerprinting · MAC address randomization

T. Mestre dos Santos (✉) · R. Neto Marinheiro · F. Brito e Abreu
Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
e-mail: tmmss1@iscte-iul.pt; rui.marinheiro@iscte-iul.pt; fba@iscte-iul.pt

© The Author(s) 2025

E. Kornyshova et al. (eds.), *Smart Life and Smart Life Engineering*,
https://doi.org/10.1007/978-3-031-75887-4_11

11.1 Introduction

The tourism sector has been growing steadily. If the pre-pandemic trend is achieved from 2023 onward, it will reach 3 billion arrivals by 2027, based on the World Bank development indicators (see Fig. 11.1).

As a consequence, the impact of tourist activities in popular destinations has risen significantly over the years, often fostered by the proliferation of cheaper local accommodation (Guttentag 2015). That increase led to exceed of carrying capacity in those destinations, a phenomenon called *tourism overcrowding*, or simply *overtourism*. The latter degrades visitors' quality of experience, reducing their feeling of safety, making it difficult to move around, enjoy the attractions, and use basic services, such as transportation and restoration, due to long wait times, while reducing the authenticity from the perspective of tourists (Tokarchuk et al. 2022). Overtourism also deteriorates the lives of local residents, due to an increase in urban noise, less effective urban cleaning, higher prices for basic goods and services (as businesses seek to capitalize on the increased demand), displacement caused by local accommodation, and cultural clashes when visitors fail to respect local customs, traditions, and privacy, sometimes leading to the former expressing negatively against the latter (Biendicho et al. 202). Last, but not least, the environmental sustainability, structures, and cultural heritage of overcrowded destinations are also jeopardized, leading to a loss of authenticity (Seraphin et al. 2018). Mitigating overtourism benefits all stakeholders:

- Local residents reduce their stress from over-occupation of personal space and privacy and improve their attitude toward tourists and tourism professionals.
- Tourism operators speed up service delivery and quality of service.

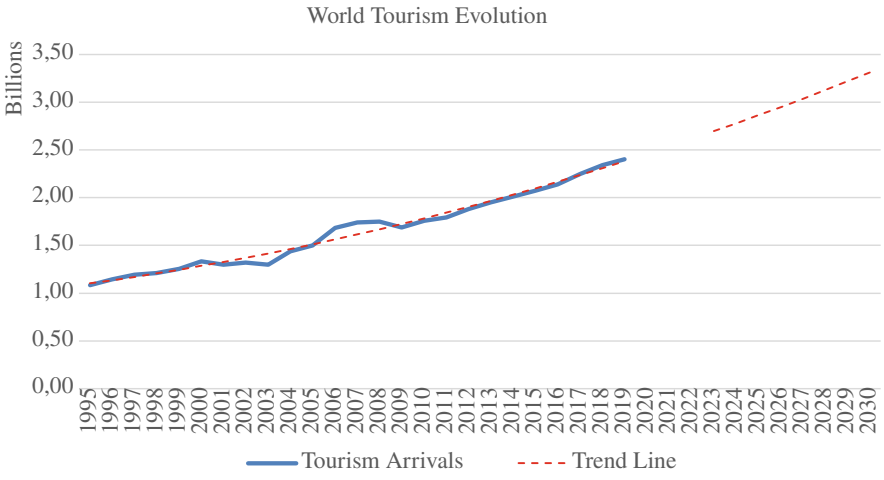

Fig. 11.1 Worldwide evolution of tourism arrivals, based on the World Bank's data

- Tourists increase their visit satisfaction, with fewer delays, increased safety, and cleanliness.
- Local authorities improve services by making just-in-time decisions and planning more effectively urban cleaning and public safety routines, as well as reducing operating costs.
- Heritage managers can prevent heritage degradation more effectively, thus retaining the authenticity of destinations.
- Local businesses increase their share of tourism income.

Overtourism mitigation actions, such as promoting the visitation to less occupied but equally attractive areas, can be applied in recreational, cultural, or religious spots, both in indoor scenarios like palaces, museums, monasteries, or cathedrals and in outdoor ones such as public parks, camping parks, concerts, fireworks, or video mapping shows.

Besides assuring a better visiting experience, those actions are also necessary for security reasons (e.g., to prevent works exhibited in a museum from deteriorating or even being vandalized by exceeding room capacity), health reasons (e.g., preventing infection in pandemic scenarios by not exceeding the maximum people density specified by health authorities), or even for resource management (e.g., to reduce the intervention of security and cleaning teams).

To implement overtourism mitigation actions, crowding information should be made available. Several approaches can be used for crowd detection, such as image capturing, sound capturing, social networks, mobile operator's data, and wireless spectrum analysis (Dias da Silva et al. 2019). The latter can be performed using passive or active sniffing methods, characterized by exploring protocol characteristics and small information breaches, such as on Wi-Fi or Bluetooth protocols, extensively used in mobile devices. Figure 11.2 provides a comparison of those approaches for crowd counting in terms of range, precision, time delay of analysis, and implementation costs.

The best option regarding cost, precision, and the near-real-time availability of data required for managing tourism crowding effectively, while complying with privacy rights, is the one based on sensing wireless communication traces, since the vast majority of tourists carry a mobile phone (Dias da Silva et al. 2019; Singh et al. 2020). Earlier approaches relied on counting the number of unique MAC (Media Access Control) addresses in messages emitted by mobile devices. However, due to user privacy concerns, most mobile devices nowadays use MAC address randomization, i.e., the same device exposes different MAC addresses over time, making it more challenging to accurately count the number of devices, thus leading to inaccurate crowd counting.

This chapter describes a low-cost approach to monitoring overtourism. It consists of a crowding sensor that performs real-time detection of trace elements generated by mobile devices from different wireless technologies, namely, Wi-Fi and Bluetooth, while addressing the MAC address randomization issue when determining the number of mobile devices in the sensors' vicinity, as an improvement over our previous work Dias da Silva et al. (2019). Another improvement refers to

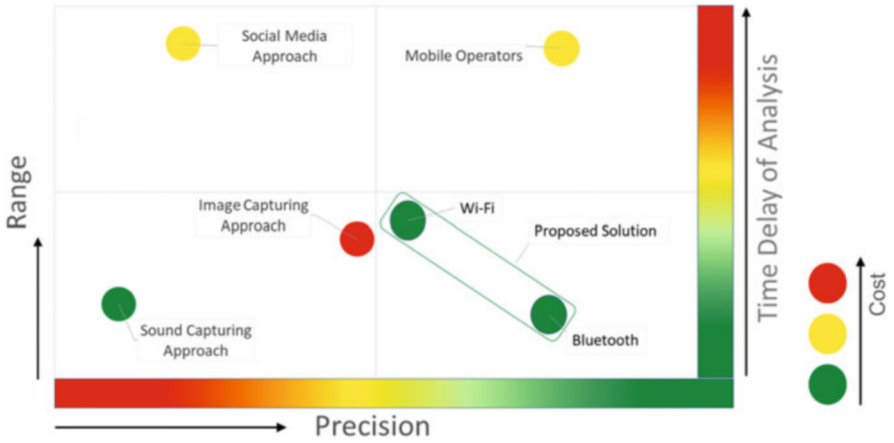

Fig. 11.2 Different approaches for crowd counting in terms of range, precision, time delay of analysis, and implementation costs (Adapted from Dias da Silva et al. 2019)

the provision of multiple communication methods for uploading the crowding information to a cloud server, by using either Wi-Fi or LoRaWAN protocols, thus mitigating network limitations on the installation location, something that has been disregarded on other approaches.

Our sensor is the basis of a Smart Tourism Toolkit (STToolkit) being built in the scope of the RESETTING[1] project, funded by the European COSME Programme., to facilitate the transition toward a more sustainable operation of tourism SMEs and improved quality of the tourism experience. The STToolkit will guide how to build and set up our sensors, either in indoor or outdoor appliances, by including support materials such as an installation manual, video tutorials, setup images, and cost calculators.

Furthermore, this research considers a correlation between the number of mobile devices and the real number of people present in an area. This assumption is especially relevant in touristic scenarios, where our sensors are aimed to be deployed since tourists usually carry their mobile phones to take pictures and record videos during their visits. Therefore, it is assumed that the number of mobile devices in a given area is directly correlated with the number of people in the same area. This is corroborated by De Meersman et al. (2016), where it is shown that mobile phone data is a valuable data source for statistical counting of people.

This chapter is organized as follows: Sect. 11.2 identifies and discusses related work; Sect. 11.3 presents the proposed architecture of a typical installation using our sensors; then, in Sect. 11.4, we describe our proposed Wi-Fi detection algorithm, which tackles the MAC address randomization issue; on Sect. 11.5, we describe

[1] RESETTING is an acronym for "Relaunching European smart and SustainablE Tourism models Through digitalization and INnovative technoloGies".

the technologies used in our solution; then, on Sect. 11.6, we present the setup and discuss the results obtained from a field validation; finally, on Sect. 11.7, we draw some conclusions and outline future work.

11.2 Related Work

Crowd counting by detecting trace elements from mobile devices' wireless activity can be performed either by the use of passive or active sniffing methods. However, only passive methods that monitor wireless traffic in a non-intrusive manner are acceptable, because active methods can cause network and user disruptions, as well as legal infringements. Many passive methods employ probe request capturing, which are messages periodically sent by devices to announce their presence to surrounding APs (Access Points), allowing a fast connection upon reaching a known network. These messages are sent in bursts and are unencrypted, meaning that they can be simply captured using passive sniffing techniques, and contain the device's MAC address. The probe request frame structure is presented in Fig. 11.3.

Earlier detection approaches relied on the device's real MAC address that was sent in the SA (Source Address) of these frames. In this case, the number of devices was simply equal to the number of different MAC addresses. However, when devices send their real MAC address, they may be easily tracked. To solve this privacy vulnerability, since 2014, manufacturers started to implement MAC address randomization on their devices. It consists of assigning to probe requests randomly generated virtual MAC addresses changing over time. Thus, the real MAC address remains unknown, protecting the user's identity and making it much more difficult to track. Unfortunately, this has led to inaccurate crowd counting and has hampered many solutions adopted until then. Moreover, the MAC address randomization process is dependent on the manufacturer and the operating system of the device, which also makes it a much more complex procedure to circumvent.

The difference between a real MAC address (globally unique) and a virtual MAC address (locally administered) is in the 7th bit of the first byte of the MAC address, as shown in Fig. 11.4. Therefore, we can simply distinguish these two types of MAC addresses by only checking this bit.

The implementation of the MAC address randomization added a level of complexity to uniquely identify devices. Therefore, the research has advanced toward the exploration of other properties and fields of the probe request frames,

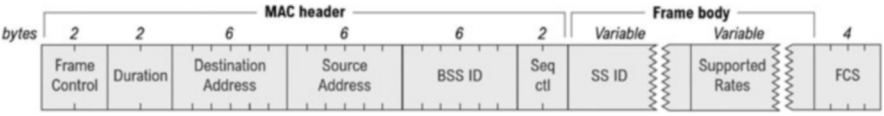

Fig. 11.3 Probe request frame (based on Institute of Electrical and Electronics Engineers)

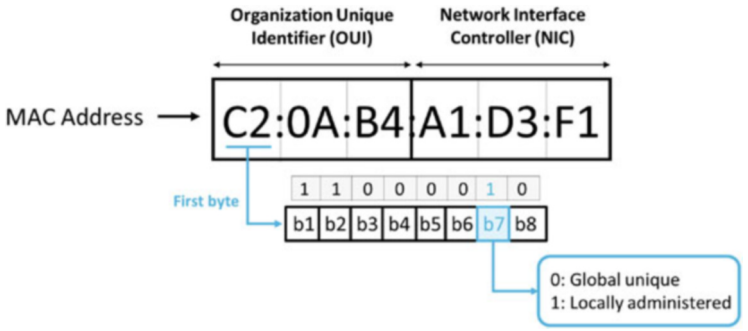

Fig. 11.4 Difference between a real and a virtual MAC address

since the MAC address is no longer a reliable option just by itself for accurate crowd counting. In spite of this, probe request frames still disclose other weaknesses that can be exploited for counting the number of devices. Several strategies have been adopted to mitigate the impact of randomization, as follows:

- **SSIDs**[2] **Comparison:** based on comparing the known networks (in terms of SSIDs) to a device, an information that is contained in the probe requests (Berenguer et al. 2022).
- **Fingerprinting:** based on generating a unique identifier (fingerprint) from other fields in probe request frames. The contents to generate this fingerprint are usually obtained from IEs (Information Elements) conveyed in the frame body of probe requests (Bravenec et al. 2022; Vega-Barbas et al. 2021).
- **Fingerprinting + Clustering:** this strategy relies not only on fingerprinting from IEs but also on other properties from these messages, such as the SEQ (SEQuence number), burst size, or IFTA (Inter-Frame Time Arrival) from probe request frames. A clustering algorithm considering these properties simultaneously can be applied, where each cluster will represent a unique device (Cai et al. 2021; Covaci 2022; He et al. 2022; Torres-Sospedra et al. 2023; Uras et al. 2020, 2022).

In Berenguer et al. (2022), the PNL (Preferred Network List), which contains the SSIDs from the known networks of a device sent in probe requests, is used for counting the number of devices in a location and distinguishing residents from visitors in the city of Alcoi in Spain. The authors claim an accuracy of 83% in detection, with some reported overestimations and incongruencies.

Most approaches rely, however, on applying fingerprinting techniques to uniquely identify mobile devices. The work reported in Vega-Barbas et al. (2021) used a network of sensors to estimate the number of persons in a given location based on IEs fingerprinting. The system was tested in public events with a considerable density of people, with a claimed accuracy close to 95%. Another

[2] SSID (Service Set IDentifier) is a sequence of characters that uniquely names a Wi-Fi network.

work described in Bravenec et al. (2022) used the same approach, considering not only the IEs but also the PNL and the recurrence of the same randomized MAC address to generate device fingerprints at a conference in Lloret de Mar, Spain, but precision is not reported.

Some other studies not only considered IEs for fingerprinting but also clustered this information along with other properties or patterns from probe requests. For this purpose, many studies used clustering algorithms that consider a combination of different features from probe requests. In Cai et al. (2021), not only probe requests but also beacons and data packets were used to count the number of devices in given locations. This method was tested with a dataset purposefully generated for the scope of this work, reaching an accuracy of 75%; however, it was not tested in a real crowded scenario. The work reported in He et al. (2022) considered IEs, SEQ, and the RSSI (Received Signal Strength Indicator) from probe requests with a neural network for estimating crowding levels in a shopping mall in Hong Kong, reaching an accuracy of slightly over 80%.

The studies reported in Uras et al. (2020, 2022) clustered fingerprinting from IEs, the incremental speed of the SEQ, the burst frequency, and the IFTA. The authors first tested the algorithm at the University of Cagliari's Campus (Uras et al. 2020), achieving an accuracy of about 91%. A follow-up to this work Uras et al. (2022) tested the algorithm first in a controlled environment, reaching an accuracy of 97%, and further inside buses in Italy for an Automatic Passenger Counting system, with a precision of 75%. Another work reported in Covaci (2022) used the same approach considering the IEs for fingerprinting and also used a clustering algorithm for combining the generated fingerprints with burst sizes and the IFTA in the canteen of the University of Twente, with an accuracy of 90%. The work described in Torres-Sospedra et al. (2023) combined fingerprints from RSSI values of Wi-Fi APs and BLE (Bluetooth Low Energy) beacons with several clustering algorithms variants for indoor positioning, achieving a precision of around 93%.

Table 11.1 summarizes the previous approaches for crowd counting, clarifying the adopted strategies to mitigate MAC address randomization and the obtained precision.

11.3 Proposed System Architecture

The proposed system architecture of the STToolkit is presented in Fig. 11.5, where sensors count the number of devices in their vicinity and periodically report the crowding information to a cloud server. The latter also has other components for making downlink communication transparent and providing uplink services for rendering the crowding information and creation of notification policies.

Table 11.1 Approaches for crowd counting, tackling MAC address randomization

Authors	Packet-type capturing	Strategies for MAC address randomization	Real scenario appliance	Precision
Berenguer et al. 2022	Probe Requests	SSIDs Comparison	Alcoi (Spain)	83%
Vega-Barbas et al. 2021	Probe Requests	Fingerprinting	Public events	90%
Bravenec et al. 2022	Probe Requests	Fingerprinting	Conference at Lloret del Mar (Spain)	Not available
Cai et al. 2021	Beacons Data packets Probe Requests	Fingerprinting + Clustering	Not available	75% (with simulated data)
He et al. 2022	Probe Requests	Fingerprinting + Clustering	Shopping mall in Hong Kong	80%
Uras et al. 2020	Probe Requests	Fingerprinting + Clustering	Campus of the Univ. of Cagliari	91%
Uras et al. 2022	Probe Requests	Fingerprinting + Clustering	Buses in Italy	75%
Covaci 2022	Probe Requests	Fingerprinting + Clustering	University of Twente campus	90%
Torres-Sospedra et al. 2023	Beacons	Fingerprinting + Clustering	Not available	93%

11.3.1 Crowding Data Collection

Each crowding sensor includes a detector responsible for passively capturing mobile devices' trace elements for each wireless technology (Wi-Fi and Bluetooth) in the sensor vicinity, an anonymized local database, where all gathered information is stored, and a detection engine, responsible for counting the number of devices by analyzing the information contained in the local database and reporting the crowding information to the cloud server. Each sensor can perform only Wi-Fi or Bluetooth detection, or both simultaneously, and can quickly switch between the technologies to be used for detection.

In this edge computing approach, data collection and crowding level measurement generation are performed locally in each sensor, so that only the number of devices detected is sent to the cloud server. So, the information to be passed is minimal, not requiring a high sampling rate for data transmission, and also protecting nodes from outside threats, since the communication line prevents the majority of attack types. Furthermore, limiting data exchange not only reduces communication costs, but also eases protection complexity for the node, and makes it easier to guarantee user privacy. Also regarding user privacy, all gathered data is anonymized before being stored in the local database.

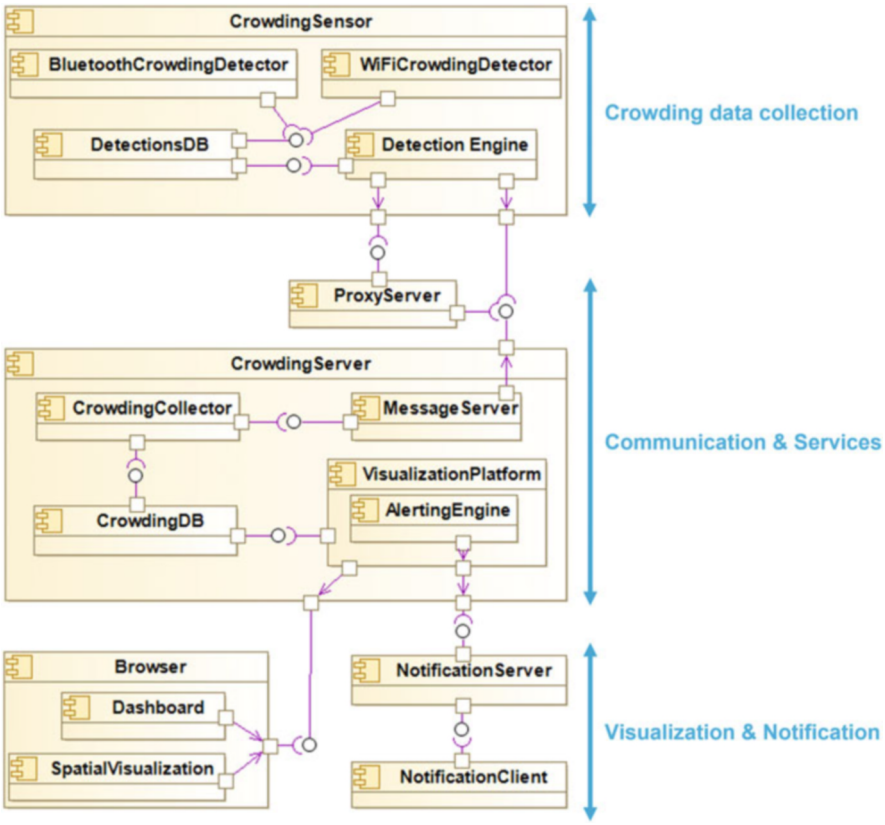

Fig. 11.5 Component diagram of the crowding detection STToolkit

11.3.2 Communication and Services

To better address installation location requirements and connectivity limitations, a flexible deployment regarding uplink technologies has been considered. Data can be uploaded to the cloud server by using a variety of communication protocols, such as Wi-Fi or LoRaWAN (Long Range Wide Area Network).

If Wi-Fi is available on-site, data can be uploaded directly to the *Message Server* via the MQTT (Message Queuing Telemetry Transport) protocol, a lightweight method of carrying out messaging, using a publish/subscribe model, widely used for IoT (Internet of Things) applications. This option can be applied straightforwardly in indoor tourism scenarios, for instance, in a museum, which generally provides a Wi-Fi network to visitors.

In outdoor scenarios, such as public parks or city squares, where overtourism situations can also arise, Wi-Fi coverage may not be available. Since sensors must upload crowding information, other approaches rely on mobile operators'

communication, which may be an expensive option, usually with monthly fees depending on the number of sensors used, each using a SIM card.

To mitigate this problem, we offer the option for uploading data via LoRaWAN, a standard of the International Telecommunication Union that provides a low-cost and scalable alternative that is feasible for our application, since sensors only communicate a small amount of data, i.e., the number of detected devices. For this, sensors must be equipped with a LoRa board and corresponding antenna to communicate the crowding information to a LoRaWAN gateway that, in turn, will route the information to the cloud server via the MQTT protocol. Regarding coverage, there are a few LoRa networks, designed for IoT appliances, that can be used for uploading data, like The Things Network open collaborative network or the, also crowdsourced, Helium network, a decentralized wireless infrastructure supported by blockchain. The Helium network adopted for this solution is the fastest growing IoT network with LoRaWAN compatibility that provides a large coverage in many countries in Europe, such as those involved in the RESETTING project. Figure 11.6 shows the Helium network coverage provided by Hotspotty in cities where our STToolkit may be deployed in the context of the RESETTING project, such as Lisbon, Barcelona, Tirana, and Heraklion, the capital of the Greek island of Crete.

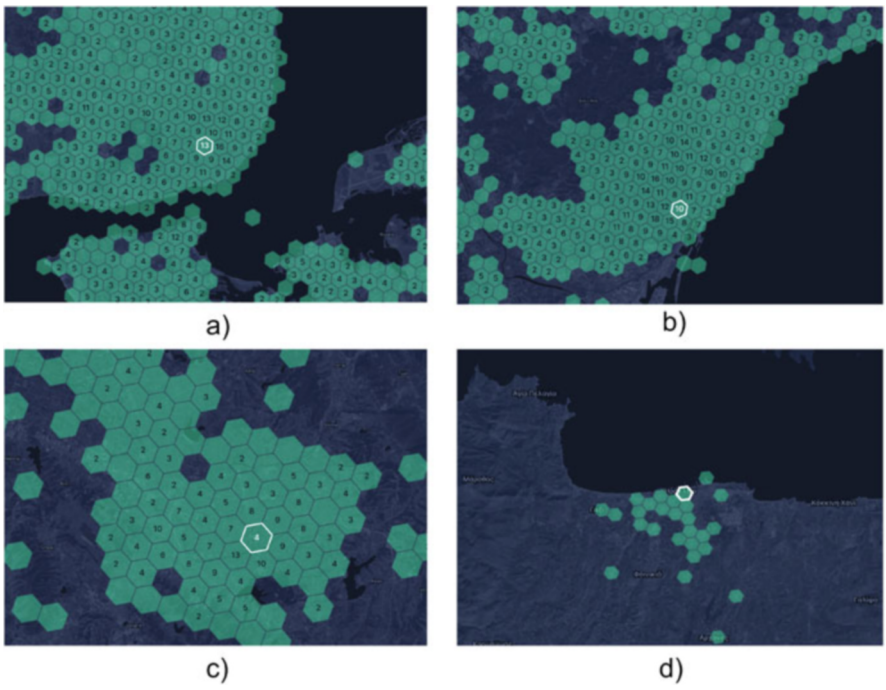

Fig. 11.6 Helium network coverage in (**a**) Lisbon, (**b**) Barcelona, (**c**) Tirana, and (**d**) Heraklion (Compiled by the authors from Hotspotty 2023)

Regarding the Helium network, an SME can choose between two alternatives for uploading crowding information: using Helium with a third-party LoRaWAN service provider, such as Helium-IoT, or using Helium with a private LoRaWAN server.

As shown in Fig. 11.5, the *Message Server* is the only entrance point for all messages in the cloud server, independently of the communication protocol used for uploading the crowding information. This provides transparency since all messages are received in the cloud server via the MQTT protocol independently of the communication technology adopted for uploading the crowding information.

Furthermore, in areas with low or no Wi-Fi or Helium network coverage, it is also possible to acquire equipment for that purpose, such as a Wi-Fi mesh system, which will allow expanding the Wi-Fi network coverage, or a Helium hotspot, for grating Helium network coverage for uploading data via the LoRaWAN protocol.

11.3.3 Visualization and Notifications

The cloud server also has several components to make downlink communication transparent and provide several uplink services that can be used by Smart Tourism Tools to understand the crowding levels in areas where each sensor is placed, with a clear and simple perspective. Possible crowding services are:

- Rendering of temporal information, as seen in Fig. 11.10
- Rendering of geographic information, as seen in Fig. 11.11
- Notification policies, e.g., when crowding threshold levels are reached
- Raw data for custom-made integrations, e.g., spatial visualization using a BIM (Building Information Model), as seen in Fig. 11.12

11.4 Proposed Wi-Fi Detection Algorithm

MAC address randomization performed by mobile device manufacturers, due to user privacy concerns, has made the identification of a mobile device a much more difficult task and, consequently, more difficult to accurately perform device counting. Therefore, an algorithm was developed for the detection of mobile devices through Wi-Fi, tackling the MAC address randomization issue using a fingerprinting technique, presented in Fig. 11.7. The explanation of each step of the proposed algorithm is presented below. A similar algorithm is also envisaged for Bluetooth detection since the randomization problem is also pertinent to this technology.

The *WiFiCrowdingDetector*, seen in Fig. 11.5, is responsible for processing each Wi-Fi packet captured by the sensor. The first operation performed is a packet-type identification (data packets, probe requests, or other type). Data packets will be used for counting the number of devices connected to an AP and probe requests for

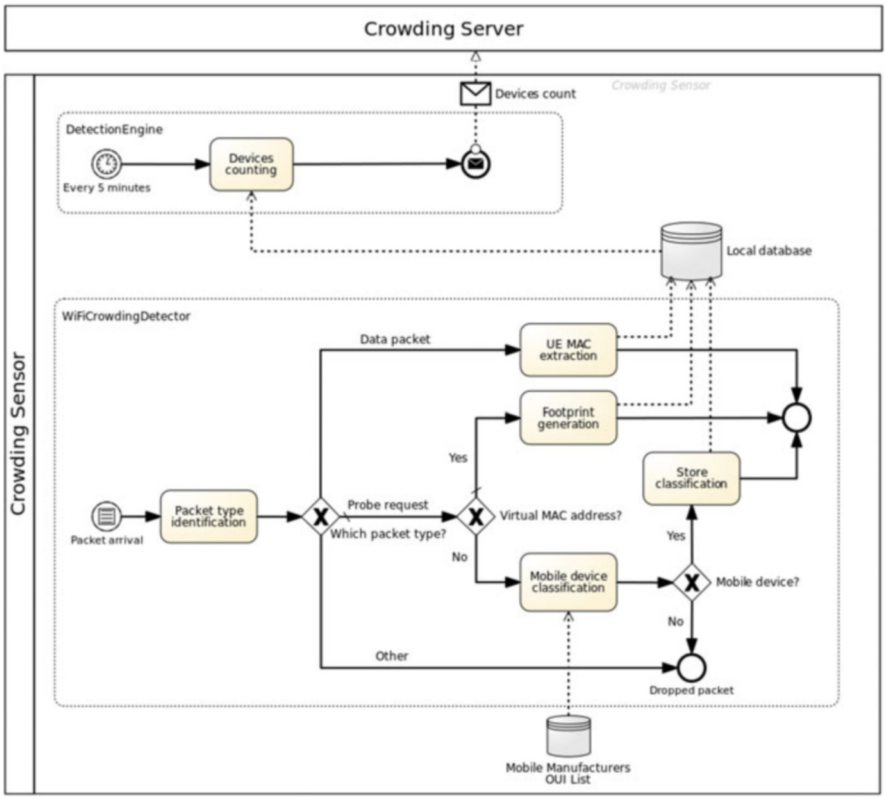

Fig. 11.7 Proposed Wi-Fi detection algorithm

counting the number of devices not connected to any AP. The packet type can be obtained by checking the Frame Control field, presented in Fig. 11.3. Packets that are not either data packets or probe requests will be immediately dropped since they are not relevant for device counting.

Regarding data packets, only the UE (User Equipment) part of the MAC address needs to be accounted for. First, it is necessary to locate it in the frame, which is performed by checking the DS (Device Status) information in the frame header, since the UE MAC address position may vary according to the direction of the frame. These MAC addresses can be directly counted as single devices because when a device is connected to an AP, the MAC address is kept constant throughout the connection and, therefore, will not change randomly. For this reason, after the UE MAC address extraction, it is directly stored in the sensor's anonymized local database.

Regarding probe requests, the first operation performed is aimed at distinguishing its Source Address between a real and a virtual MAC address. This is performed by checking the 7th less significant bit of the 1st octet of the MAC address, as already

Table 11.2 Probe request's information elements used to create device fingerprint

Information element	IE ID	IE length	Description
Supported rates	1	<8	Data transfer rates supported by the device
Extended supported rates	50	<256	Other bit rates supported by the device
DS parameter set	3	1	Device's channel setting when sending a probe request
HT capabilities	45	26	Compatibility with the 802.11n standard
VHT capabilities	191	12	Compatibility with the 802.11ac standard
Extended capabilities	127	<256	Other device capabilities
RM-enabled capabilities	70	5	Information for measuring radio resources
Interworking	107	<9	Interworking service capabilities of the client
Vendor specific	221	<256	Vendor-specific information (e.g., device manufacturer)

shown in Fig. 11.4. To follow the trace of a real MAC address, a device classification is applied, aimed at only counting MAC addresses that belong to mobile devices. This is done by checking the address's OUI (Organizational Unique Identifier):[3] if the OUI matches one of the known mobile manufacturers, obtained from the Wireshark manufacturer database, the MAC address should be considered as a mobile device, and it must be counted and stored in the local database; otherwise, the MAC address is not considered as a mobile device and is discarded.

To follow the trace of a virtual MAC address, a fingerprinting technique must be performed to uniquely identify devices that use MAC address randomization. For this, the IEs contained in the frame body of the probe request are analyzed. For each IE, the entirety of its information is considered, including the IE ID, Length, and Value bytes. For those IEs with substantially varying values across probes emitted from the same device (e.g., DS Parameter Set), only the bytes of IE ID and Length are analyzed. Table 11.2 shows the IEs used for the fingerprinting technique. After analyzing all IEs, a hash function is applied to all its contents. As a result, a 64-bit footprint is generated for each probe request and stored in the local anonymized database. Then, all the devices that are trying to connect to a Wi-Fi network, by sending probe requests to discover available networks in proximity with a virtual MAC address, are uniquely identified by the footprint. So, each footprint should be counted as one mobile device using MAC Address randomization that is trying to connect to a Wi-Fi network. To avoid counting the same device twice, if the

[3] OUI is a part of the MAC address identifying the network adapter vendor.

same MAC address is captured in both data packets and probe requests, it is only accounted for once.

Then, the *Detection Engine* will periodically count the number of devices detected within a sliding window of X minutes, i.e., the number of devices detected in the last X minutes, and upload that information to the cloud server. Both the sliding window period and data sampling rate can be independently and easily configured by the user. Since we intend to provide real-time or near-real-time data availability, the data sampling rate of our sensors needs to comply with this requirement. So, we have chosen a 5-minute period for the data sampling rate, as it is a sufficient time period for providing near-real-time data availability. Also, the same 5-minute period was chosen for the sliding window, so that each crowding measurement could comprise all detected devices within each sliding window.

The number of devices detected is the sum of (i) the number of devices connected to an AP, obtained from the UE MAC addresses captured in data packets, plus (ii) the number of different devices not connected to any AP, obtained from the MAC addresses from probe requests with real MAC addresses, plus (iii) the number of different footprints from probe requests with virtual MAC addresses.

11.5 Adopted Technologies

The developed STToolkit uses a variety of open-source software technologies, installed in off-the-shelf hardware available at affordable costs. Figure 11.8 presents the UML deployment diagram proposed for the STToolkit concerning all technologies adopted in our solution.

For the operating system of our sensors, we have opted for a Kali Linux distribution, which comes with a large number of preinstalled network tools that can be easily used for detecting devices in different technologies. For the local database, a SQLite database was chosen for storing all gathered data, which requires low memory usage, while meeting all other requirements. For data anonymization, the sensors use the t1ha library that provides several terraced and fast hash functions. In particular, we have opted for the t1ha0 hash function, as it is one of the fastest available at the library.

To perform Wi-Fi detection, the required hardware is a Wi-Fi card that supports monitor mode, which allows the board to capture all network traffic in its proximity. We have chosen the Alfa Network AWU036AC board for our sensor, which provides high performance at a low cost, having two antennas for dual-band detection (2.4 GHz and 5 GHz) without interfering with Bluetooth devices. As for the sniffing software, we have chosen the Aircrack-ng tool, an open-source software with several different applications for detecting devices. In particular, we use *airmon-ng* for enabling the monitor mode in the Wi-Fi board, and *airodump-ng* for capturing raw Wi-Fi frames. For Bluetooth detection, we have selected the Ubertooth-One board and corresponding BlueZ package that contains tools and frameworks for Bluetooth usage in Linux.

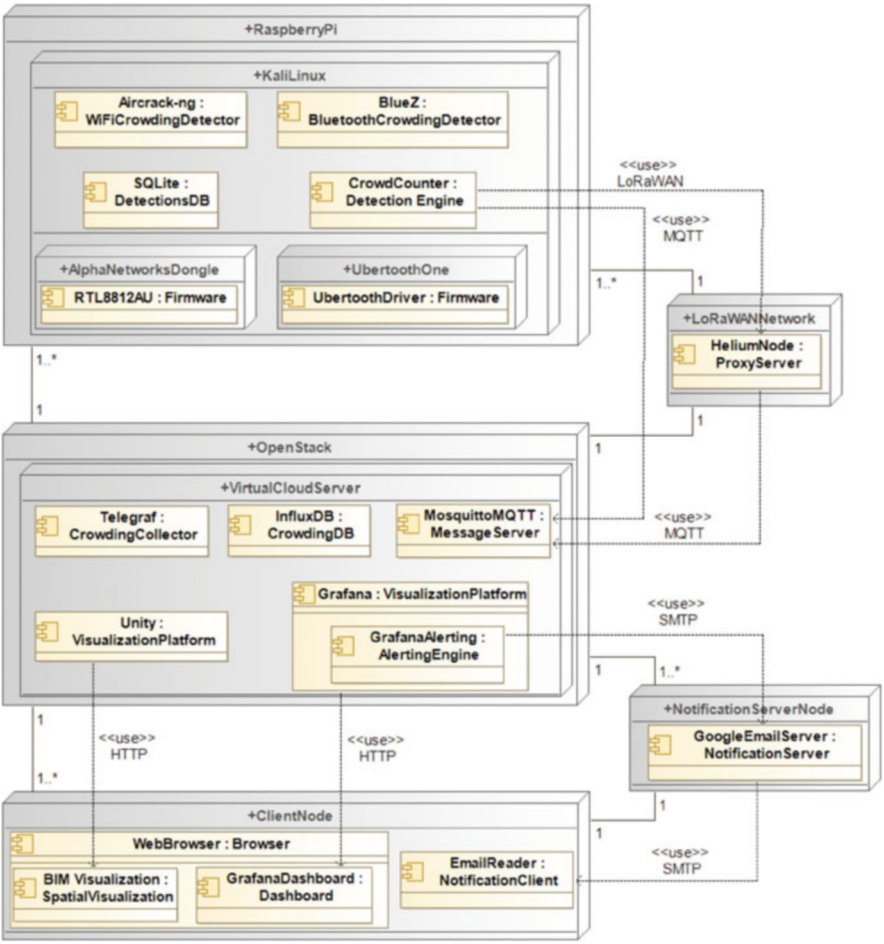

Fig. 11.8 Deployment diagram for the crowding detection STToolkit

For receiving all messages, our cloud server uses the Mosquitto MQTT, a lightweight message broker that implements the MQTT protocol.

For the data ingestion of all measurements sent by sensors, a database is necessary. This database has to be lightweight, capable of querying data rapidly from timestamps, and also capable of providing support for data visualization platforms to observe the results in real-time. That is why we chose InfluxDB, a time-series database focused on IoT applications, for our *CrowdingDB* component. For the *CrowdingCollector*, responsible for pushing all messages received in the *Message Server* via MQTT protocol to the *CrowdingDB* in the appropriate format, we chose Telegraf, an open-source plugin-driven server agent for collecting and reporting metrics from devices.

Table 11.3 Prototype hardware components and respective functions

Component	Function
Raspberry Pi 3/4	Coordinate and process
Alfa Network AWUS036AC	Wi-Fi detection
Ubertooth-One	Bluetooth detection
Raspberry Pi IoT LoRa pHAT	Upload via LoRaWAN (if necessary)

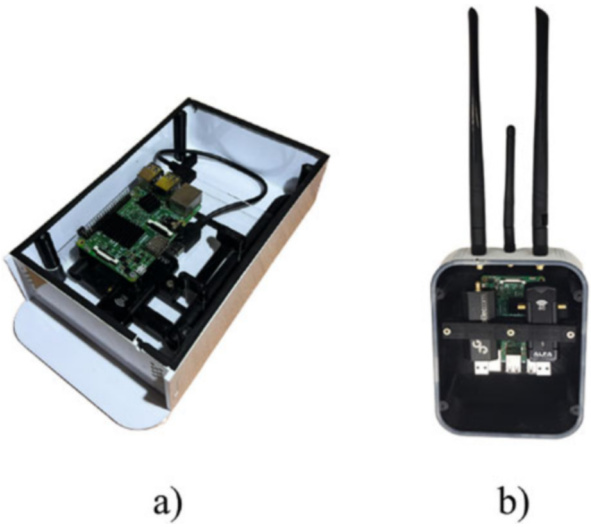

Fig. 11.9 Sensor cases: (**a**) large version with no exposed antennas; (**b**) small version with exposed antennas

Finally, for data visualization, we chose Grafana, an open-source analytics and monitoring tool compatible with several databases, including *InfluxDB*. This framework can be used for creating custom dashboards with graphs and panels for viewing, with different spatiotemporal levels of granularity, the crowding information. Additionally, *Grafana* can be used for creating notification policies, allowing users to receive alerts according to the crowding levels via a diversity of contact points.

A prototype was developed whose hardware components and respective functions are illustrated in Table 11.3. The prototype uses custom-designed cases adapted to deployment locations, either with exposed antennas or not, as shown in Fig. 11.9. These prototype versions have been deployed at several locations at our university campus to test the operation and performance of the STToolkit, which is further described in the next section.

Furthermore, as our sensor's processing unit is a single computer board, namely, a Raspberry Pi, there are multiple options for powering our sensor, either directly from a battery or even via USB ports or Power over Ethernet (PoE), even though the most straightforward and convenient alternative should be to directly connect it to a mains power supply through a transformer.

11.6 Field Validation and Discussion

The prototype described in the preceding section was designed and implemented to withstand all the scenarios where the sensors may be deployed.

To test and validate the STToolkit architecture, sensors have been placed at several spots across Iscte's campus, and crowding information has been collected since September 2022.

The sensors were deployed both indoors and outdoors, in places with different crowding patterns, such as areas with a large pedestrian flow, internal and external passages between buildings, and places for prolonged stays, such as a large study hall and the university library.

This field experiment has been conducted with the sole purpose of assessing the perception of the crowding phenomena in the university campus, rather than the accuracy regarding the real number of people at each location. It focused on perceiving crowding patterns and tendencies, such as time breaks between classes, lunch periods, and highly populated events. The aim was to assess how sensors could perceive relative variations throughout the days across the several locations of the campus and how quickly the sensors were able to detect them.

The accuracy was addressed in other contexts in a more controlled environment (Santos 2023), where the detections from sensors were compared with the real number of people, obtained through direct observation during a public event, to assess the effectiveness of the solution.

The crowding data has been used for visualization, using a variety of temporal dashboards, and maps that highlight the geographic distribution of crowding at each location where the sensors have been deployed. Data has also been used for spatial visualization in the form of heatmaps and also, for a more realistic view, using avatars on top of Iscte's BIM (Building Information Model).

Dashboards allow users to select time ranges for crowding data temporal visualization, to perceive people's concentration and flows during specified periods, and to identify highly populated events. Figure 11.10 shows a comparison of crowding levels during a normal day of classes at Iscte's campus, at the selected spots where sensors have been deployed.

In addition to the temporal rendering of the information, data has also been used for spatial visualization in the form of heatmaps, to grant users a better perception of people distribution at several locations where the sensors are deployed. This can be seen in Fig. 11.11, where it is possible to perceive the crowding hotspots from our sensors deployed at Iscte's campus at a given time.

Moreover, raw crowding information can also be easily used by third-party integrations. To validate this, we built a walking avatar animation upon Iscte's BIM, to achieve a more realistic perception of space occupancy, as shown in Fig. 11.12, for one of the campus buildings. There, the number of detected devices, obtained in real time from sensors, determines the number of ingress and egress avatars in their areas of detection. This last experience was performed during the International

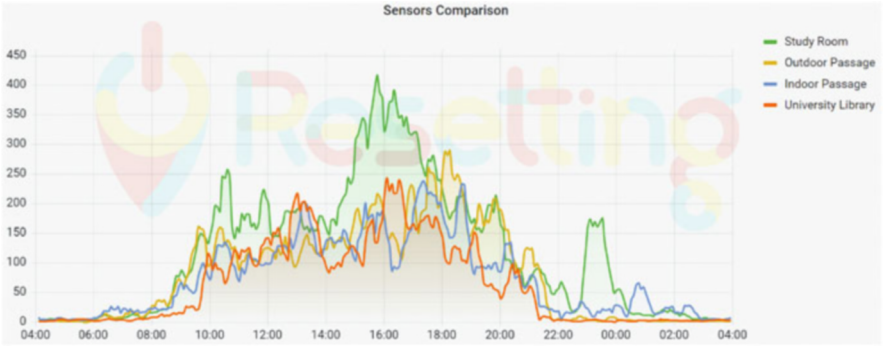

Fig. 11.10 Comparing crowding at Iscte's campus

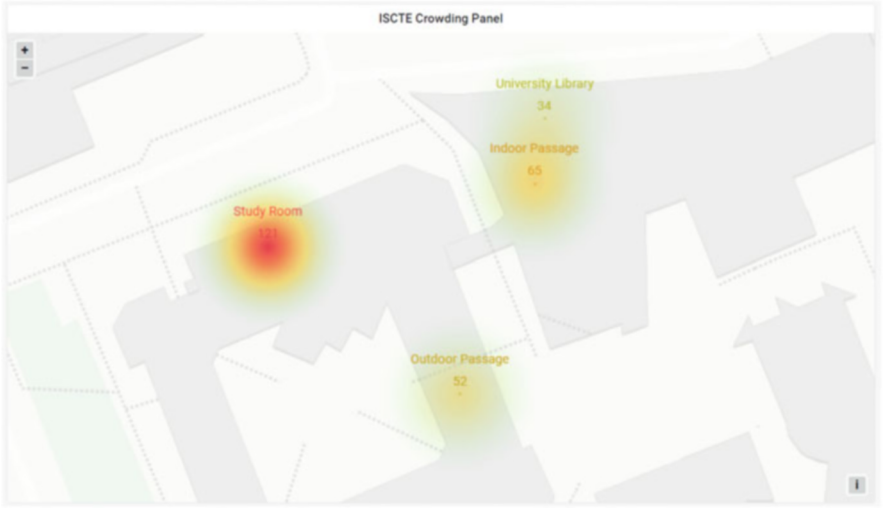

Fig. 11.11 Crowding hotspots at Iscte's campus

Posters & Demos Workshop on Smart Tourism held by the RESETTING project at Iscte in January 2023.

Furthermore, it is also possible to create notification policies, where alerts can be triggered if predetermined crowding thresholds are exceeded, by using several contact points such as email, Telegram, Google Chat, Microsoft Teams, Slack, or PaperDuty, enabling users to make just-in-time decisions facing overtourism situations. These alerts can be easily configurable by using the *Grafana* tool, also used for spatiotemporal visualization of crowding information.

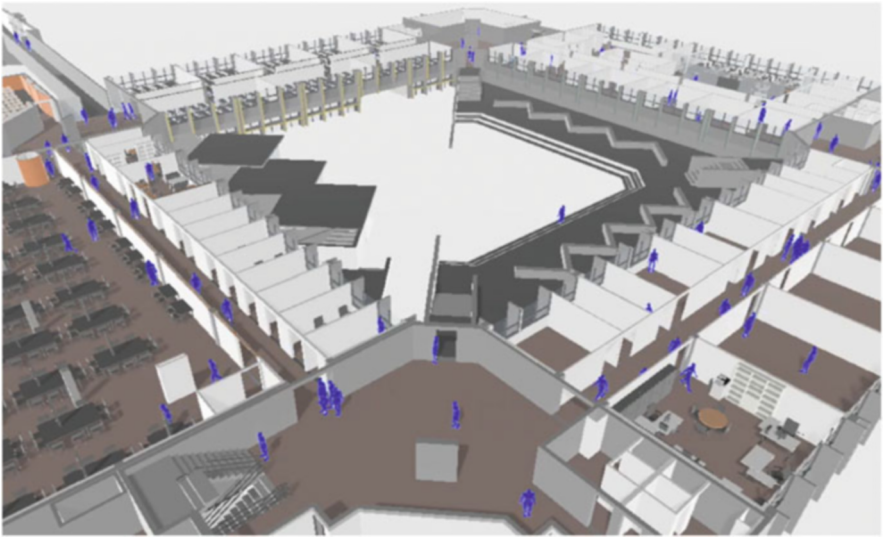

Fig. 11.12 Crowding visualization based on avatars upon the campus BIM

11.7 Conclusions and Future Work

Overtourism deteriorates the visiting experience of tourists, the quality of life of residents, as well as the environment. By monitoring it, tourism professionals can identify areas of concern and put measures in place to lessen its negative effects, encouraging better tourism practices to ensure that tourism benefits both tourists and locals while preserving natural and heritage resources.

For monitoring overtourism, a low-cost approach based on mobile devices' activity has been developed. The sensors, equipped with off-the-shelf hardware available at affordable costs, perform real-time detection of trace elements of mobile devices' wireless activity, mitigating MAC address randomization, and crowding values are put together in a cloud server. Alternative communication channels for uploading the crowding information, namely, via Wi-Fi or LoRaWAN protocols, allow for addressing local connectivity limitations at the installation location of sensors. In addition, scalability is provided by maintaining the hardware costs low, by using open-source software, and by the simplicity of the installation and configuration of each sensor. Regarding the RESETTING project, an SME must choose the option that best fits its needs and requirements for implementing its system. To help with this purpose, the STToolkit will include a sensor deployment calculator for SMEs to estimate the most cost-effective uplink alternative according to the installation location of each sensor.

The crowding information can then be analyzed by destination managers to understand the crowding levels in areas where each sensor is placed in a clear and simple perspective, either by dashboards for temporal or spatial visualization of crowding information or using the raw data for custom-made integrations.

Furthermore, notification policies can be created when overtourism situations occur, opening the possibility of implementing just-in-time mitigation actions required by the nature of these circumstances, as they may be sudden and unpredictable.

Preliminary tests have been conducted for this solution. A prototype version of the crowding sensor was deployed at several spots on Iscte's campus, in typical usage scenarios with a high flow and/or extended presence of people, such as the university library, a large study hall, and two passageways. Crowding information has been collected and used to monitor people's flow and detect high-crowding events on campus.

In the short term, this STToolkit will be deployed at the Pena Palace, one of the most iconic tourism sites in Portugal, surrounded by a large walkable park, flagellated by overtourism all year round. The objectives will be promoting alternative routes for tourists within the park, limiting their number in sensitive areas, and making the tourism offer in this area more sustainable. Our detection approach will then contribute to reducing the overwhelming feeling of pressure in critical hotspots, thus leading to a greater visiting experience for tourists who visit this attractive tourist site.

Furthermore, a second prototype version of sensors is also envisaged. The latter will include new boards with greater performance, new antennas with higher gains for larger detection ranges, directional antennas for performing detection in specific areas, custom-designed heatsinks for the processing units to achieve the best possible performance, as well as new custom-designed cases.

Further details on the sensors, including demos of setting up and configuring the edge nodes and the cloud server, can be found online at the RESETTING@Iscte site.

Acknowledgments This work has been developed in the scope of the RESETTING project, funded by the European COSME Programme (EISMEA), under grant agreement COS-TOURINN 101038190. The cloud-based infrastructure (computing and storage) used was provided by the INCD, Funded by FCT and FEDER under project 01/SAICT/2016 nº022153. The current work has also been supported by Fundação para a Ciência e Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through national funds and, when applicable, co-funded by European Union (EU) funds under the project UIDB/EEA/50008/2020.

References

Berenguer, A., Ros, D.F., Gómez-Oliva, A., Ivars-Baidal, J.A., Jara, A.J., Laborda, J., Mazón, J.N., Perles, A.: Crowd monitoring in smart destinations based on GDPR-ready opportunistic RF scanning and classification of WiFi devices to identify and classify visitors' origins. Electronics **11**(6), 835 (2022). https://doi.org/10.3390/electronics11060835

Biendicho, M., Papaoikonomou, E., Setó-Pamies, D.: Tourists go home! Examining antitourism in Barcelona from an emotions perspective. Tour. Culture Commun. **22**(3), 275–295 (2022). https://doi.org/10.3727/109830421x16345418234010

Bravenec, T., Torres-Sospedra, J., Gould, M., Fryza, T.: What your wearable devices revealed about you and possibilities of non-cooperative 802.11 presence detection during your last IPIN visit.

In: 2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE, Piscataway (2022). https://doi.org/10.1109/ipin54987.2022.9918134

Cai, Y., Tsukada, M., Ochiai, H., Esaki, H.: MAC address randomization tolerant crowd monitoring system using Wi-Fi packets. In: Asian Internet Engineering Conference. ACM, New York (2021). https://doi.org/10.1145/3497777.3498547

Covaci, A.I.: Wi-Fi MAC address randomization vs crowd monitoring. Tech. rep., University of Twente (2022). http://essay.utwente.nl/91744/

De Meersman, F., Seynaeve, G., Debusschere, M., Lusyne, P., Dewitte, P., Baeyens, Y., Wirthmann, A., Demunter, C., Reis, F., Reuter, H.I.: Assessing the quality of mobile phone data as a source of statistics. In: European Conference on Quality in Official Statistics, pp. 1–16 (2016)

Dias da Silva, R., Neto Marinheiro, R., Brito e Abreu, F.: Crowding detection combining trace elements from heterogeneous wireless technologies. In: 2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC). IEEE, Piscataway (2019). https://doi.org/10.1109/wpmc48795.2019.9096131

Guttentag, D.: Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. Curr. Issues Tour. **18**(12), 1192–1217 (2015). https://doi.org/10.1080/13683500.2013.827159

He, T., Tan, J., Chan, S.H.G.: Self-supervised association of Wi-Fi probe requests under MAC address randomization. IEEE Trans. Mobile Comput. 1–14 (2022). https://doi.org/10.1109/tmc.2022.3205924

Hotspotty: Helium hotspot map (2023). https://explorer.helium.com

Institute of Electrical and Electronics Engineers: IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. https://doi.org/10.1109/ieeestd.2021.9363693

Santos, T.M.: Smart tourism toolkit for crowd-monitoring solutions. Master's thesis, Iscte - Instituto Universitário de Lisboa (2023). http://hdl.handle.net/10071/29505

Seraphin, H., Sheeran, P., Pilato, M.: Over-tourism and the fall of Venice as a destination. J. Destination Marketing Manag. **9**, 374–376 (2018). https://doi.org/10.1016/j.jdmm.2018.01.011

Singh, U., Determe, J.F., Horlin, F., Doncker, P.D.: Crowd monitoring: state-of-the-art and future directions. IETE Tech. Rev. **38**(6), 578–594 (2020). https://doi.org/10.1080/02564602.2020.1803152

Tokarchuk, O., Barr, J.C., Cozzio, C.: How much is too much? Estimating tourism carrying capacity in urban context using sentiment analysis. Tour. Manag. **91**, 104522 (2022). https://doi.org/10.1016/j.tourman.2022.104522

Torres-Sospedra, J., Quezada Gaibor, D.P., Nurmi, J., Koucheryavy, Y., Lohan, E.S., Huerta, J.: Scalable and efficient clustering for fingerprint-based positioning. IEEE Internet Things J. **10**(4), 3484–3499 (2023). https://doi.org/10.1109/JIOT.2022.3230913

Uras, M., Cossu, R., Ferrara, E., Bagdasar, O., Liotta, A., Atzori, L.: WiFi probes sniffing: an Artificial Intelligence based approach for MAC addresses de-randomization. In: 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, Piscataway (2020). https://doi.org/10.1109/camad50429.2020.9209257

Uras, M., Ferrara, E., Cossu, R., Liotta, A., Atzori, L.: MAC address de-randomization for Wi-Fi device counting: combining temporal- and content-based fingerprints. Comput. Netw. **218**, 109393 (2022). https://doi.org/10.1016/j.comnet.2022.109393

Vega-Barbas, M., Álvarez-Campana, M., Rivera, D., Sanz, M., Berrocal, J.: AFOROS: a low-cost Wi-Fi-based monitoring system for estimating occupancy of public spaces. Sensors **21**(11), 3863 (2021). https://doi.org/10.3390/s21113863

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.