CrowdMap: Spatiotemporal Visualization of Anonymous Occupancy Data for Pandemic Response

Sitao Min[1], Ritesh Ahuja[2], Yingzhe Liu[2], Abbas Zaidi[2],
Catherine Phu[2], Luciano Nocera[2], Cyrus Shahabi[2]
sm2370@rutgers.business.edu,{riteshah,yingzhe,abbaszai,chphu,nocera,shahabi}@usc.edu
[1]Rutgers Business School — Rutgers NJ, USA
[2]University of Southern California — Los Angeles CA, USA

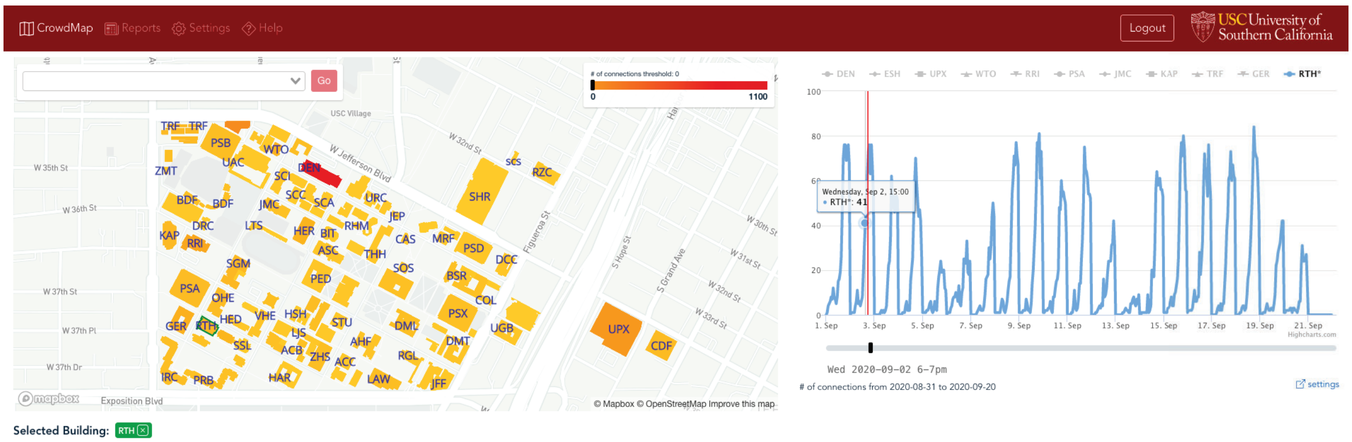

Figure 1: CrowdMap front-end showing daily building aggregates of active connection counts on the USC campus

ABSTRACT

CrowdMap is an anonymous occupancy monitoring system developed in response to the COVID-19 pandemic. CrowdMap collects, cleans, and visualizes occupancy data derived from connection logs generated by large arrays of Wi-Fi access points. Thus, CrowdMap is a passive digital tracking tool that can be used to reopen buildings safely, as it helps actively manage occupancy limits and identify utilization trends at scale. Occupancy monitoring is possible at various levels of resolution over large spatial (e.g., from individual rooms to entire buildings) and temporal (e.g., from hours to months) extents. The CrowdMap web-based front-end implements powerful spatiotemporal querying and visualization tools to quickly and effectively explore occupancy patterns throughout large campuses. We will demonstrate CrowdMap and its spatiotemporal GUI that was deployed for an entire university campus with data continuously being collected since summer 2020.

CCS CONCEPTS

• **Information systems** → **Spatial-temporal systems**; • **Human-centered computing** → *Information visualization.*

KEYWORDS

Spatio-temporal Occupancy Data, Visualization, Wi-Fi Monitoring

ACM Reference Format:
Sitao Min[1], Ritesh Ahuja[2], Yingzhe Liu[2], Abbas Zaidi[2],, Catherine Phu[2], Luciano Nocera[2], Cyrus Shahabi[2]. 2021. CrowdMap: Spatiotemporal Visualization of Anonymous Occupancy Data for Pandemic Response. In *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21), November 2–5, 2021, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474717.3484269

1 INTRODUCTION

Recording the number of occupants in a building that may be exposed to close encounters has become vital to managing COVID-19 transmission and complying with local and government social distancing policies. In principle, occupancy measurement is straightforward to conceive, but in practice, is difficult to implement. A number of solutions that can be easily deployed in a small business/office setting exist. These include check-in systems that people use to log their locations and dwell time (either by manual input in an online form or by scanning a code at the entrance to the building) and self-report positive tests to assist in contact tracing [1]. These systems however have seen limited adoption [6, 13] as they are intrusive — requiring every individual's diligent cooperation and willingness to volunteer their information — and are difficult to scale and enforce in organizations with extended working spaces such as large businesses, universities, industry campuses, and hospitals.

To reconcile these challenges, digital contact tracing tools, such as mobile apps on GPS-equipped devices have been extensively

SIGSPATIAL '21, November 2–5, 2021, Beijing, China

ACM ISBN 978-1-4503-8664-7/21/11.
https://doi.org/10.1145/3474717.3484269

deployed throughout Asia and Europe [4, 7, 8]. App-based contact tracing has the potential to address manual contact tracing's limitations of scalability, notification delays, memory recall challenges, and contact identification in public spaces. However, using GPS to detect close encounters in buildings is not viable since GPS signals are blocked or reflected by walls. Location privacy concerns in addition to their potential for mass surveillance and data abuse have led to unexpectedly low adoption of COVID-19 contact tracing apps. This can be seen with state-sponsored systems [11, 15] and universities, for example at the University of Southern California (USC), we were unable to convince decision makers to use even a privacy-focused app [17].

Instead, at USC, we have developed **CrowdMap**, an anonymous occupancy system designed to actively manage occupancy limits, reopen buildings safely, and identify utilization trends at scale. CrowdMap uses non-intrusive passive digital tracking by leveraging the university Wi-Fi network management system logs. Highly accurate occupancy data is typically obtained via dedicated hardware such as cameras, infrared sensors, and magnetic reed switches, but retrofitting large campuses with such devices is expensive and would ultimately raise its own set of privacy concerns [12]. A less intrusive solution for indoor workplace occupancy monitoring relies on Wi-Fi as devices are automatically set to connect to the access point transmitting the strongest signal, which can then be associated with the office or room where the person is located. Similar methods have been studied in isolation to solve the indoor localization problem [5], with application to context-aware and location-based services [18, 19]. Wi-Fi triangulation is often used to derive the position of connecting devices with accuracy whenever multiple routers with varying strength are *visible* to the device. However, we do not use triangulation in CrowdMap as it can be used for tracking individuals. Our demonstrated solution is amenable to large organizations that rely on a *network management system* to monitor Wi-Fi internet access and optimize user experience. These systems routinely monitor connecting devices at the level of individual Wi-Fi access points (AP) and log the connection information including the device MAC address and connection period, and a user identifier if on a secured network.

Due to the presence of stationary devices that also utilize the same network, CrowdMap employs sophisticated algorithms to derive the number of individuals connected. We report only aggregate occupancy statistics processed from de-identified AP logs in a privacy-conscious fashion. As shown in Figure 4, to optimize Wi-Fi signal for users most buildings are outfitted with several APs per floor, and spaces with large capacities (e.g., classrooms, labs, athletic training facilities) often have multiple APs. We released CrowdMap in Spring 2021 to authorized USC users and have been collecting occupancy data starting in August 2020. We demonstrate CrowdMap on this data to show (i) how we ingest and process connection logs to derive good estimates of occupancy counts, and (ii) how end-users utilize its spatial and temporal querying and visual analytics capability. The GUI allows connecting spatial and temporal information of university campuses at different resolutions together, thus powering a reliable and effective way to inform campus re-opening, and related policy decisions in a pandemic such as COVID-19.

2 CROWDMAP SYSTEM

Physical Infrastructure. The assets involved in enabling the proposed passive monitoring system include thousands of Access Points (AP) connected to a centralized controller and a sophisticated network operations system (NOS) running on dedicated servers. USC uses the Aruba Airwave network management system [16], but similar systems include Cisco DNA Center [2] and Juniper's Mist [3] to name a few. The APs are strategically placed throughout USC campuses in order to optimize Wi-Fi signals to users, and as a byproduct, are ideal for passively capturing the whereabouts of individuals on campus. Data packets transmitted in existing Wi-Fi traffic are analyzed by the network system for the received signal strength (RSS) and the MAC address of each connected device. Since these devices must periodically transmit data to maintain their association with the access point, location tracking of these devices is conducted by the NOS through simple signal strength measurements taken by multiple APs that see the tracked client. This allows the APs to record the approximate position of the device in 3D space around the signal transmitter.

Network Connection Logs. Aruba Airwave logs are generated nightly at each of the three USC campuses: the University Park Campus (UPC), its adjacent North University Park (NUP), and the Health Sciences Campus (HSC). Information on all connecting devices during the 24 hour period is logged internally by the system. Prior to being uploaded to the CrowdMap server, logs are de-identified by replacing the device MAC address with an MD5 hash. Received logs are in the form of compressed CSV files where each row corresponds to a connection event (a device connecting or disconnecting from an AP) and includes the following fields: the anonymized device identifier (MAC address), the anonymized USC account username (if not available, this column is set to the device identifier), the wireless AP name, the connection start date-time, the disconnection date-time, the connection duration in minutes and the wireless network name (a unique Service Set Identifier). To derive occupancy counts from logs additional processing is needed as detailed in the next paragraph.

Hourly Occupancy Counts. CrowdMap tracks occupancy through hourly counts estimates computed at each AP by post-processing the network connection logs in three filtering stages. In the first stage, we merge subsequent short connections from the same device to the same AP (e.g., one connection starts within 1 minute after another one ends). In the second stage, connections generated by remote or static devices (e.g., workstations, printers, sensors) are filtered out based on thresholds that were set empirically by site-wide data analysis; we reject 'short' connections with a duration of less than 5 minutes, and 'long' connections with a duration of more than 12 hours. In the third stage, devices that connect to the same AP multiple times during an hour are consolidated into a single record. To ease their use the resulting hourly occupancy counts are stored in JSON format.

Hourly occupancy counts can be generated in three ways: (i) on-demand by the CrowdMap web services, (ii) on a fixed schedule, using a CRON job running the CrowdMap Command Line Interface (CLI) or (iii) by manually batch processing log files with the CLI.

Aggregate occupancy statistics are generated on the fly from the hourly counts per AP. Aggregation over time, such as daily,

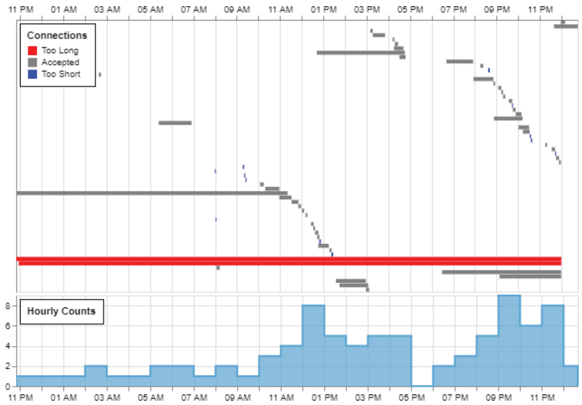

Figure 2: Example AP logs and hourly counts for one day

weekly, or monthly occupancy counts and aggregation over space, such as floor and building occupancy counts are produced by the Crowdmap services when requested by the front end according to user preferences.

Tables 1 and 2 present descriptive statistics on connection logs and occupancy counts, respectively, for a period of nine months from August 18, 2020 to May 30, 2021. The connection logs show an average of 931,725 daily connections from 19,899 connected devices, 76.9% of which were pruned away as noise. Of the rejected connections, 97.8% were too short and 2.2% were too long. Upon further analysis, we saw that the 'short' connections are often generated by older generation mobile devices that attempt to connect to an AP with outdated wireless networking standards. The occupancy counts after filtering and aggregation are reported in Table 2. Whereas, Figure 2 focuses on a single AP over a 24 hour period on May 30, 2021.

Building Data. Building data includes building metadata (e.g., building name, full name, address), their geographical extents (i.e., footprints) and detailed floor plans that include the exact position of all APs on the floors. Building footprints are used as a base layer for visualizing occupancy counts, floor plans are used to overlay AP location and occupancy counts. Pre-existing footprints were obtained from Los Angeles GeoHub [10], and footprints of recently constructed buildings were created manually using a 'WYSIWYG' online editor [9]. Building metadata, floor plans, and AP information were obtained from the Aruba Airwave system. Overall, the three USC campuses (UPC, HSC, NUP) include 297 buildings, 7459 APs, and 797 floors.

System Architecture We follow a tiered architecture having data, services, and presentation layers, as shown in Figure 3. The data layer is implemented as flat files including daily connection logs, hourly occupancy counts, and user preferences. The services layer is implemented in Node.js which includes entry points to services for: (i) occupancy data at different levels of temporal aggregation

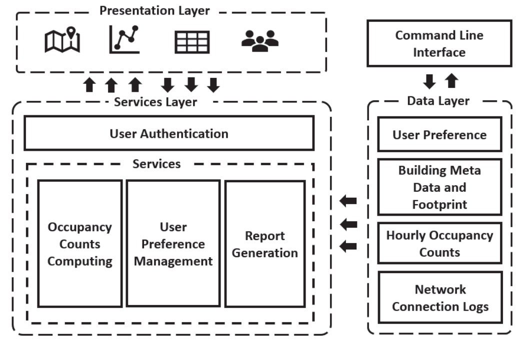

Figure 3: System Architecture

(hourly, daily, weekly , and monthly), (ii) building metadata and footprints, (iii) flat files listing; (iv) user preferences, and (v) report generation (leveraging LINQ [14] to index and query). The presentation layer consists of a Web-based front-end built with Vue.js and includes an interactive map built with Mapbox.gl and several charts built with HighCharts and Plotly.

Deployment We deployed CrowdMap on a production-ready system provisioned by USC IT Services and secured access using the Single Sign-On (SSO) solution used at USC. User access is controlled via the SSO dashboard while the user preferences are managed from within the application.

3 INTERACTIVE VISUALIZATION TOOLS

These responsive and interactive visual tools allow users to explore occupancy data for spatial regions and temporal ranges of interest.

Buildings choropleth map. We implemented an interactive choropleth map that overlays aggregated building occupancy counts over the building footprints, as shown in Figure 1 (left). A time slider allows the user to see in real-time how these counts vary over time on the map. Users can visually filter buildings with high occupancy by setting a lower bound occupancy threshold (slider and associated color scale in the legend of the map of Figure 1).

Floor plans dot maps. These custom interactive visualizations are made available when selecting buildings on the map, and display AP and occupancy counts on top of building and floor maps. An example of such a map is shown in Figure 4. We implemented floor plans dot maps using D3.js, with the floor plan images obtained from the Aruba Airwave system. The dot map consists of red dots showing the AP locations with corresponding occupancy counts shown as blue bubbles. When hovering the mouse over a bubble we display a tooltip showing AP information and occupancy counts. Buttons let users navigate different floors and a time slider allows to configure the time for which the occupancy counts are shown.

Time series line charts. To enable users to study occupancy trends at scale and over time, we have implemented various time series line charts using Highcharts.js. Line charts are generated

# of log files	855	**Period covered**	08/20	to	05/21
Total # of devices	328k	**Daily avg. # of devices**	19k		
Connection logs count	266m	**Daily avg. # of connections**	931k		
Avg. # of rejected logs per day 347,401 (76.9%) (**long**: 97.8%, **short**: 2.2%)					
Avg. connection duration per log: 38.6 mins					

Table 1: Network Connection Logs Data Statistics

	Per access point	Per floor	Per building
Hourly counts	0.39	2.95	11.94
Daily counts	3.93	31.01	121.43
Weekly counts	27.40	209.36	847.06

Table 2: Occupancy Data Statistics

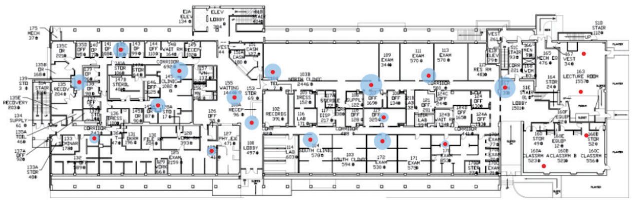

Figure 4: Floor plans dot map with bubble chart of AP counts

for different aggregation levels based on users' time period preferences, and show temporal trends for buildings, floors, and access points. To ease interpretation, line charts showing building occupancy aggregates are displayed next to the buildings choropleth map (see Figure 1, right), and line charts showing floor and AP level occupancy are shown next to the floor plan dots maps.

Configurable reports. These reports complement the exploratory features of the GUI. The reports UI allows users to specify a desired time period and any subset of buildings. Reports include the detailed occupancy counts statistics (similar to data shown in Tables 1 and 2) and plots implemented using plotly.js.

4 SPATIAL/TEMPORAL QUERYING.

CrowdMap offers advanced features to enable its user to do free-form exploration over multiple spatial and temporal resolutions, providing quick and effective access to recent occupancy counts and utilization trends at scale. Our powerful querying and visualization tools aim to provide a reliable and effective way to obtain data to inform policy decisions, manage occupancy limits, and reopen buildings safely.

Spatial exploration. Spatial exploration is powered by the bird's eye view of the building choropleth map and complemented by the floor plan dots map to further resolve occupancy information at building, floor, or access point granularity. Interactivity is a key design choice in CrowdMap. For instance, when a user clicks on a building on the map, floor plans dot maps of selected buildings are automatically populated on the interface. We also provide a search box where users can search buildings by name.

Temporal exploration. A settings page allows users to configure an arbitrary date range at which to resolve the occupancy data. This mechanism allows the user to adjust the time resolution of the occupancy data to visually analyze short and long-term trends. When the period specified in the settings is less than three weeks, the user is presented with a view of hourly AP counts. If the period is more than three weeks and less than six weeks, weekly counts are shown, and if the period is greater than six weeks, monthly aggregates are shown. All the visualizations are dynamically adapted to the aggregation level according to the period specified by the user settings. This mechanism allows users to view occupancy information of each building at a micro time level e.g. what period of the day or in which day of the week there is the most number of people, and also to view occupancy information at a macro time level, e.g. which week of the semester there were more occupants in the building.

5 DEMONSTRATION AND ARTIFACTS

For the demonstration, we will showcase various aspects of our CrowdMap system to the conference participants, including the connections log data, data pre-processing pipeline, collected data statistics, and the specific aspects of user interface, with the spatial and temporal visual analytics. We will first demonstrate how occupancy data are generated from connection logs and how the data is visualized in the front-end depending on the user preferences. We will then show how users interact with the dashboard to explore the data in space and time for their buildings of interest, e.g., for ongoing occupancy monitoring and reporting. Finally, we will create a sandbox of our CrowdMap dashboard including some sample data that conference participants will be able to access. Participants will be able to act as users to examine connection counts trends, monitor for compliance and create reports.

Acknowledgements. We thank USC Information Technology Services for providing us with anonymous aggregate Wi-Fi Access Point data. This research has been funded in part by NSF grants IIS-1910950 and CNS-2027794, and funds from the USC's VP of Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] [n.d.]. *Back to Campus - Contact Tracing with SafeZone.* Retrieved May 24, 2021 from https://www.criticalarc.com/contact-tracing-with-safezone/
[2] [n.d.]. Cisco DNA Center. https://developer.cisco.com/dnacenter/.
[3] [n.d.]. Juniper Networks powered by Mist AI. https://www.mist.com/resources/video-virtualize-indoor-location-mist/. Accessed: 2021-06-05.
[4] [n.d.]. USC Trojan Check. https://keepteaching.usc.edu/faculty/full-toolkit/campus-health-wellness/trojancheck/. Accessed: 2021-06-05.
[5] Moustafa Abbas, Moustafa Elhamshary, Hamada Rizk, Marwan Torki, and Moustafa Youssef. 2019. WiDeep: WiFi-based accurate and robust indoor localization system using deep learning. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE, 1–10.
[6] Natalie Fedor. 2020. *Purdue officials delve into contact tracing, data dashboard details.* Retrieved May 24, 2021 from https://www.purdueexponent.org/campus/article_f22a0944-cc5a-11ea-a6d8-d7bd25d2ecf7.html
[7] Wei Jiang, Lianjie Shu, Honghao Zhao, and Kwok-Leung Tsui. 2013. CUSUM procedures for health care surveillance. *Quality and Reliability Engineering International* 29, 6 (2013), 883–897.
[8] Marcel Jonker, Esther de Bekker-Grob, Jorien Veldwijk, Lucas Goossens, Sterre Bour, and Maureen Rutten-Van Mölken. 2020. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. *JMIR mHealth and uHealth* 8, 10 (2020), e20741.
[9] OpenStreetMap Mapbox. [n.d.]. geojson.io. https://geojson.io/.
[10] City of Los Angeles Hub. [n.d.]. City of Los Angeles Hub. https://geohub.lacity.org/. Last accessed May 29, 2021.
[11] Emily Seto, Priyanka Challa, and Patrick Ware. 2021. Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness. *Journal of Medical Internet Research* 23, 3 (2021).
[12] Cyrus Shahabi, Seon Ho Kim, Luciano Nocera, Giorgos Constantinou, Ying Lu, Yinghao Cai, Gérard G Medioni, Ramakant Nevatia, and Farnoush Banaei Kashani. 2014. Janus-Multi Source Event Detection and Collection System for Effective Surveillance of Criminal Activity. *JIPS* 10, 1 (2014), 1–22.
[13] Samantha Subin. 2021. *These college students are working as contact tracers to stop the spread of Covid on campus.* Retrieved May 24, 2021 from https://www.cnbc.com/2021/03/11/stopping-covid-on-campus-college-students-work-as-contract-tracers.html
[14] Mads Torgersen. 2007. Querying in C#: How Language Integrated Query (LINQ) Works. In *ACM SIGPLAN OOPSLA*.
[15] Séverine Toussaert. 2021. Upping uptake of COVID contact tracing apps. *Nature Human Behaviour* 5, 2 (2021), 183–184.
[16] Aruba Airwave Network Management website. Last accessed May 29, 2021. https://www.arubanetworks.com/products/network-management-operations/airwave.
[17] Li Xiong, Cyrus Shahabi, Yanan Da, Ritesh Ahuja, Vicki Hertzberg, Lance Waller, Xiaoqian Jiang, and Amy Franklin. 2020. REACT: real-time contact tracing and risk monitoring using privacy-enhanced mobile tracking. *SIGSPATIAL* (2020).
[18] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas Spanos. 2017. Freedetector: Device-free occupancy detection with commodity wifi. In *IEEE (SECON Workshops)*.
[19] Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J Spanos. 2018. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. *Energy and Buildings* 174 (2018), 309–322.