1-2020

Practical server-side indoor localization: Tackling cardinality outlier challenges

Anuradha RAVI
Singapore Management University, anuradhar@smu.edu.sg

Archan MISRA
Singapore Management University, archanm@smu.edu.sg

Citation

RAVI, Anuradha and MISRA, Archan. Practical server-side indoor localization: Tackling cardinality outlier challenges. (2020). *2020 International Conference on COMmunication Systems & NETworkS COMSNET: January 7-11, Bengaluru, India: Proceedings*. 192-199.
Available at: https://ink.library.smu.edu.sg/sis_research/4871

Practical Server-side Indoor Localization: Tackling Cardinality & Outlier Challenges

Anuradha Ravi
School of Information Systems
Singapore Management University
anuradhar@smu.edu.sg

Archan Misra
School of Information Systems
Singapore Management University
archanm@smu.edu.sg

***Abstract*—In spite of many advances in indoor localization techniques, practical implementation of robust device-independent, server-side Wi-Fi localization (i.e., without any active participation of client devices) remains a challenge. This work utilizes an operationally-deployed Wi-Fi based indoor location infrastructure, based on the classical RADAR algorithm, to tackle two such practical challenges: (a) low cardinality, whereby only the associated AP generates sufficient RSSI reports and (b) outlier identification, which requires explicit identification of mobile clients that are attached to the Wi-Fi network but outside the fingerprinted region. To tackle the low-cardinality problem, we present a technique that uses cardinality changes to demarcate periods of stationary behaviour, and then augment the RSSI reports with useful but apparently "stale" RSSI readings from neighbouring APs. To tackle the filtering of clients with outlier locations, we propose a model that combines a weighted path-loss propagation model with a Voronoi tessellation of the fingerprint map to define suitable boundary values for RSSI readings. We experimentally show how these two approaches improve the stability and robustness of location tracking, and consequently, the accuracy of overall occupancy estimation.**

***Index Terms*—Location based services, WLAN network measurements,**

I. Introduction

Wi-Fi based location tracking is one of the most popular and widely-researched indoor localization techniques. Multiple techniques such as fingerprinting [1], propagation modeling [2] and time-of-flight estimation [3] have been proposed, with state-of-the-art techniques reportedly achieving location errors less than 10 cm. However, most such localization schemes require *active client participation* (often requiring an App on the mobile device) and thus have adoption challenges for universal coverage. There are several commercial applications of indoor location analytics, such as occupancy counting [4] and visitor movement analytics [5], that ideally require the retrieval of indoor location tracking of *every Wi-Fi enabled* device in a *passive* manner–i.e., without their explicit participation. Our work in this paper is motivated by our broader, ongoing work on accurate and timely *occupancy sensing* ("how many individuals/devices are presently at a specific indoor location?"), as part of a building-block for occupancy-aware, energy-efficient building operations.

Server-side localization approaches (e.g., [6]) help address this challenge, as the infrastructure essentially uses measurements of regular device transmissions to automatically localize them without their explicit cooperation. Recent work on server-side localization explains [6] how commercial Wi-Fi deployments support such server-side localization by providing real-time feeds (so-called RTLS messages) of such signal strength measurements (RSSI) from multiple APs. Such coarse-grained RTLS data are processed by *fingerprinting-based, nearest-neighbor* approaches (such as RADAR [7]) to estimate the actual location. However, there are two challenges that we empirically observed with such fingerprinting-based, server-side solutions:

- *Low Cardinality:* As detailed in [8], to conserve energy, modern mobile nodes (*MN*s) generate Wi-Fi PROBE_REQUESTs fairly infrequently (especially when stationary), reducing the number of distinct APs that provide fresh RSSI measurements of a mobile device. This low cardinality problem is even more acute for 802.11a (5.5GHz) bands, where neighboring APs rarely overhear transmissions due to the larger number of non-overlapping channels. The resulting location estimates are often made using readings from just one AP, resulting in median localization error rates in excess of 6-8 meters.
- *Inclusion of Extraneous Outliers:* This problem is not as well-documented. The actual localization techniques (e.g., nearest neighbor or K-nearest neighbor) implicitly assume that a mobile device is always within the fingerprinted region, and typically *snap* the unknown location of a device to the nearest pre-established *landmark* or the centroid of multiple *nearby* landmarks. We shall show that, as a consequence, MNs that attach to the Wi-Fi network but are outside the fingerprinted region (e.g., users connected to a building Wi-Fi while being in the parking lot outside) get incorrectly mapped to one of the fingerprinted landmarks, effectively polluting the occupancy estimate.

Key Contributions: We shall quantify the two challenges mentioned above and then present techniques to reduce the resulting localization errors. We make three key contributions:

- *Use Older RTLS readings to Enhance Cardinality:* To overcome the localization error that comes from primarily using the reports from just a single AP, we propose to additionally include the *apparently stale* readings reported by other APs (to which the mobile device is not presently associated). Such stale readings are an indirect manifestation of an absence of PROBE_REQUESTs from an MN, which itself usually implies that the MN is stationary. PROBE_REQUEST scans, across multiple channels, are usually triggered whenever an MN actually moves, at which point it is very likely for a *new AP* to generate an RTLS report for this MN. Accordingly, the *appearance of a new AP in the "reporting set" is a good indicator of the MN's mobility*. More importantly, during the intervening stationary periods, we can utilize even stale reports from other APs (generated after the prior movement interval), as their *RSSI readings are very likely to still be valid*.

- *Eliminate Outlier Clients:* To eliminate clients outside the fingerprinted area of a building, we first compute the Voronoi tesselation for each landmark (in the RSSI space), denoting the set of RSSI tuples that get snapped to a specific landmark. We then use a propagation model (whose parameters are derived from the fingerprint readings) to *estimate* the likely RSSI values (for each AP) at multiple *boundary points* of the fingerprinted region. Finally, we use such boundary RSSI values to define more-restricted Voronoi regions–i.e., for every individual landmark, we define a per-AP RSSI threshold representing the lowest acceptable signal value for a location that is both legitimate (within the fingerprinted area) *and* should get mapped to this landmark. MNs whose readings lie outside this threshold for *any one* of the reporting APs are then marked as outliers, and are then iteratively mapped to other alternative locations, until eventually being filtered out from the occupancy estimates.
- *Demonstrate Real-world Performance Gains:* We utilize the server-side location technology deployed at two university locations (a research lab within an academic building and the common spaces of a residential building) to demonstrate the results. In particular, we show that our "augmentation by curated stale reports" approach increases the average cardinality of useful AP reports from 1 to at least 2-3, thereby reducing the localization error from $\sim$8-10meters to $\sim$ 2-4 meters. Additionally, the outlier elimination approach also significantly reduces the error in occupancy estimation by over 80%.

We emphasize both the *generalizability* and *practical impact* of our work: (a) the challenges of low cardinality & outlier elimination apply not just to RADAR but any other passive localization technique (e.g., based on propagation models), and (b) indoor occupancy analytics is projected to be a USD $9B market by 2025[1]. The rest of the paper is organised as follows. Section II outlines relevant prior work, while Section III discusses the experimental setup and data collection. Section IV details the two problems of specific interest. Section V details the proposed solutions, while Section VI presents the experimental results. Finally, Section VII concludes the paper.

[1]https://www.marketwatch.com/press-release/indoor-location-by-positioning-systems-market-size-to-surpass-447-cagr-up-to-2025-2019-04-29

II. Related Work

In this section we present the existing approaches towards Wi-Fi based indoor localization, as well as describe their associated challenges. The Wi-Fi fingerprinting-based approach for indoor localization was introduced classically via techniques such as RADAR [7] & Horus [9] that utilize coarse-grained RSSI information. In more recent years, a variety of approaches (e.g., PinLoc [2]) have used additional physical-layer information (e.g., the phase & amplitude of different sub-carrier frequencies) to enrich the fingerprint. Alternately, ArrayTrack [10] provides accurate ($\sim$20cm) location estimates by coordinating the AoA estimates from multiple cooperating APs.

Researchers have also extensively investigated the sensitivity of RSSI-based fingerprinting strategies to various parameters. Xia et a. [11] studied how the average localization error was inversely proportional to the number of landmarks. Kaemarungsi et al. [12] showed that RSSI values in indoor environments are prone to fluctuation due to changes in ambient conditions and *shadowing* effects caused by the presence of humans. Mazuelas et al [13] proposed a robust indoor positioning system by dynamically calibrating the propagation model using the obtained real-time RSSI values. Talvitie et al [14] discussed the impact of missing fingerprint and compare techniques to interpolate and extrapolate fingerprinted values, while Kafrawy et al [15] have developed indoor WiFi propagation models for localizing vehicles and humans. To overcome the RSSI variations due to client heterogeneity, Kjaergaard [16] proposed a hyperbolic fingerprinting method that uses ratios of signal strengths (instead of absolute values) for localization.

Overall, all of these approaches implicitly assume the availability of a set of RSSI measurements from multiple APs, likely obtained through *active scans issued by an MN*–as we shall see, **current WiFi clients rarely perform such active scans**. These methods also focus on improving an MN's location accuracy within the fingerprinting area, instead of explicitly tackling the problem of filtering out MNs that attach to an AP but are outside the fingerprinted region.

III. Experimental SetUp and Data Collection

Our work utilizes the server-side indoor location system [6] that has been deployed in our university campus since 2013. The main campus Wi-Fi network utilizes an Aruba infrastructure. For the analysis in this paper, we restricted ourselves to analyzing a research lab (roughly $400m^2$ in area) located on one floor of an academic building. To demonstrate *vendor-independence*, we additionally extended this core server-side based system to a new university-owned, off-campus residential building that uses the Cisco infrastructure. Figure 1 illustrate the floor plan of the research centre.

A. Fingerprinting Process

As is conventional practice, the initial fingerprinting mechanism involves the enumeration of landmarks followed by a manual process of collecting network RSSI measurements

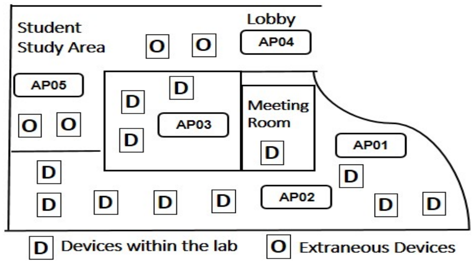

Fig. 1. Floor Plan of Research Lab (Academic Building)

at these specific landmarks. In our deployment, landmarks are delineated based on the overhead sprinklers, which are typically deployed in a grid-like fashion with a separation of 3 meters. For the Aruba infrastructure, the Real-time Location System (RTLS) mechanism caused each AP to push RSSI reports for all attached client devices[2], once every 5 seconds. For the Cisco infrastructure, we receive the RSSI feeds from Meraki Cloud [17]. As the data received from Meraki Cloud arrives aperiodically and in random order, the system waits until it receives new reports from all APs before proceeding on to localization, resulting in location update periods of $\approx 1 - 1.5$mins. Table I summarizes the key data fields received in such RTLS reports, for both the Aruba and Cisco deployments.

TABLE I
ACCESS POINT DATA AS RECEIVED FROM ARUBA AND CISCO MERAKI

Field	Description	Aruba	Cisco
Timestamp	AP Epoch Time (Milliseconds)	Yes	Yes
Client MAC	SHA1 of original MAC address	Yes	Yes
Age	Elapsed seconds since AP received a device's frames	Yes	No
Channel	2.4/5GHz band on which device was seen by AP	Yes	No
AP MAC	APs MAC Address	Yes	Yes
Associate Status	Association of a device with particular AP	Yes	Yes
RSSI	Signal Strength for a device as reported by AP	Yes	Yes

The localization process consists of two phases described below:

- **Offline Phase:** In the offline (fingerprinting) phase, we collect the RSSI reports for MNs that were deliberately spaced at each individual landmark. Based on the RADAR algorithm, for each landmark, we compute the mean of the multiple RSSI readings from each AP, and thereby create an $N - dimensional$ "landmark vector" (N=the number of APs, K=the number of landmarks). The eventual fingerprint database then consists of such landmark vectors for all of the predefined landmarks, represented as follows:

$$< L_1 :< AP_1, RSSI_1 > ... < AP_n, RSSI_n > ... \\ L_k :< AP_1, RSSI_1 > ... < AP_n, RSSI_n >> \quad (1)$$

- **Online Phase:** In this operational phase, we retrieve the RTLS feeds generated by the infrastructure (reported every 5secs for Aruba and approx. every 1.5 minutes for Cisco). The resulting files contain multiple entries for the same MN (client), with the fields as specified in Table I. *Note that the Aruba infrastructure explicitly provides an "age" field indicating the time elapsed since the last measurement by the AP*–in prior work [8], this field has been used to filter out the *stale* readings. After consolidating all the entries for a single MN, the RSSI-distance (in signal strength space) is computed across all MNs, using Equation 2, where AP_{Rn} represents the real time RSSI value reported by the AP, and AP_{Fn} represents the fingerprinted RSSI value.

$$\sqrt{(AP_{R1} - AP_{F1})^2 + ... + (AP_{Rn} - AP_{Fn})^2} \quad (2)$$

B. Experimental Evaluation

To understand the performance issues of the baseline algorithm, we manually placed mobile devices at different locations (at varying distances from one or more landmarks). We then used the server-side RSSI readings to first compute the estimated landmark location. For a given AP_i, let $RSSI_i(F, L)$ denote the fingerprinted RSSI value (at landmark L) and $RSSI_i(R)$ represent the measured RSSI value for the test client. Intuitively, the closer the MN's location to landmark L, the smaller should be the difference between $RSSI(R)$ and $RSSI_i(F, L)$. By varying the MN's location, we can obtain different values for this difference: $RSSI{-}Diff(L) = |RSSI_i(F, L) - RSSI(R)|$.

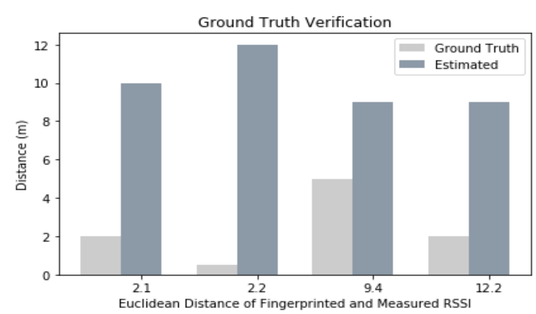

Fig. 2. Ground Truth vs Estimated Distance

Figure 2 plots both (a) the true physical distance between a measurement location l and its nearest landmark L_{near}, referred to as *Ground Truth* distance), and (b) the distance between l and the estimated landmark L_{est}; the x-axis represents the"RSSI-distance" between the measurement at l and the fingerprinted value at L_{est}. Across multiple experiments, we see that the two distances are *significantly different*, implying

[2]While RTLS reports are generated for both associated and non-associated devices, we restrict ourselves only to associated clients for our current analysis as analysis of non-associated devices has to additionally contend with the "*MAC address randomization*" problem.

that MN is incorrectly assigned to a landmark that is quite distinct and far from L_{near}. Interestingly, such errors happen for a wide range of "RSSI distances".

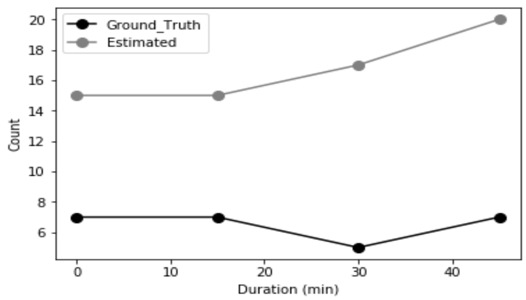

Fig. 3. Occupancy Count - Expected vs Estimated

Similarly, Figure 3 plots the observed vs. ground truth *occupancy count* of individual devices within the Residential Building, over an observation duration of 45 minutes. We see that the occupancy count is consistently *over-estimated*, providing strong evidence that multiple devices located outside the fingerprinted common spaces of the building are being erroneously localized to these common spaces. We shall discuss the reasons behind these large estimation errors in Section IV.

IV. Detailing the Identified Problems

The analysis performed on the data collected from multiple locations reveal several problems, of which we focus on two in this paper.

A. Low Cardinality

To get connected with Wi-Fi network, mobile devices broadcast PROBE_REQUEST messages, which are received by all APs in its vicinity. Typically, the AP with the highest signal strength is then chosen for association, with all data frames subsequently sent to the chosen AP. MNs typically transmit such probe requests either in the 2.4 GHz or the 5GHz band, but based on the network configuration, the APs either reply with the probe response in the same band as requested by the device, or in the preferred 5GHz band. (In modern networks, the 5.5GHz band is selected preferentially, due to the larger number of channels and the reduced interference.) For devices which do not support the 5GHz band, the APs send probe response in 2.4GHz space [17].

From our experimental data, we observed two distinct patterns of RSSI reports generated by an AP (corroborating the phenomena reported in [8]):

- The chosen AP (i.e., the one to which an MN is associated) effectively refreshes the reported RSSI values reported at each update, as it continually engages in data packet exchanges with the MN.
- In contrast, the other APs report the MN's RSSI value only during the *probing* phase, when the MN actively scans across all channels; during the subsequent data transfer phase, they do not receive any packets from the MN on their usual operating frequency.

This observation in line with the Aruba [18] specifications citing that the "associated AP" reports the RSSI values for clients based on the data frames, while the un-associated APs report RSSI values only based on any received PROBE_REQUEST packets. Literature [19] also suggests that a client send probe requests with greater frequency in the 2.4GHz space, or when it does not see any "known SSIDs" in its vicinity.

Due to these reasons, only the associated AP sends updated RSSI values, while the other APs simply report stale readings (observed during the probe phase). Table II illustrates one such case, which shows the RSSI values reported by all the APs for a particular client. The age parameter is low only for the associated AP (AP02), while the other non-associated APs ($status = 0$) report the same (stale) RSSI value (which prior approaches [8] filter out), albeit with increasing *age*.

TABLE II
Data Reported For Client-ABC in the research centre

Timestamp	Age	AP MAC	RSSI	Associated Status	Client MAC
533	33	AP01	-74	0	ABC
535	2	AP02	-60	BSSID of Associated AP	ABC
536	33	AP03	-65	0	ABC
536	34	AP05	-61	0	ABC

B. Outlier Location

Most APs have a fairly long range–in our university, we can hear APs that are 60 meters away. Accordingly in Figure 1, the range of AP01 and AP03 can extend well beyond the boundaries of the research lab, even percolating to the next building. For devices located within the fingerprinted building, this range is a positive–it effectively increases the cardinality of RSSI reports. However, this also comes with a drawback: devices can attach the Wi-Fi network from locations outside the fingerprinted region (e.g., from outside the building) and can then erroneously be localized within the fingerprinted area (as current localization techniques do not incorporate any explicit outlier elimination logic).

To illustarte this phenomenon, Table III presents an exemplar of RSSI readings, reported by 3 different APs for a case where the MN in question is actually located outside the building. The Table plots both the true SNR (higher the SNR, stronger the signal) value reported for the MN (column *Reported SNR*), as well as the fingerprinted SNR values corresponding to the landmark L (column $Fingerprint(L)$ to which the MN is currently localized. We note the significant difference between these two RSSI values (Equation 2) for each of the APs, which is indicative of a potential localization error.

V. Proposed Solution

In this section, we present our proposed solutions for the problems defined in the previous section.

TABLE III
DATA FOR A CLIENT LOCALIZED TO LANDMARK L IN THE RESIDENTIAL BUILDING

Access Points	Actual Reported SNR	Fingerprint SNR Value (L)
AP01	12	30.4
AP02	24	10.6
AP03	9	37.6

A. Low Cardinality

Prior studies [20], [21] have shown that the decline in probe requests, and the consequent drop in AP cardinality, occurs for stationary clients; in contrast, when a client moves, it effectively initiates a probing stage, which helps multiple APs to obtain RSSI estimates for the MN. Accordingly, we hypothesize that the appearance of a newly reporting AP is a marker of significant movement by an MN.

Our proposed solution exploits this phenomenon–we effectively first classify if a device is in the *moving/probe vs. stationary* stage, by observing whether there are new APs (other than the currently associated AP) that have generated new RSSI values. If so, this is likely due to the explicit PROBE_REQUEST scans initiated by a moving MN; in this case, we use all the *recent* AP reports, discarding all *stale* reports (those with timestamps older than 15 secs). However, if there are no new reporting APs, then we conclude that the MN is still *stationary* and then include even the older (stale) RSSI readings, as those readings are likely to be persistent for a stationary device. Algorithm 1 outlines the relevant pseudocode.

Algorithm 1 Stationary Pattern Identification

for $R :< RTLS >$ **do**
  if $R.AP \neq ArchivalTable.AP$ **then**
    Add $RTLS_{AP}$ to $< AP_{List} >$ for Location Estimation
    $NewFound$=1 //Here, New Found indicates a new AP association $\rightarrow$ client is moving.
  end if
end for
for $R : RTLS$ **do**
  if $ArchivalTable_{AP} = R.AP$ **then**
    if $R.Age \leq Archival\_Table\_AP.age$ **then**
      Add RTLS AP to $< AP_{List} >$ for Location Estimation
    else if $RTLSAge \geq ArchivalTableAge$ & $NewFound \neq 1$ **then**
      Add R record to $< AP_{List} >$ for Location Estimation //Include stale AP data.
    else if $R.Age \geq ArchivalTableAge$ & $NewFound = 1$ **then**
      Do Not Add R to $< AP_{List} >$ for Location Estimation //new AP found $\rightarrow$ ignore stale AP data.
    end if
  end if
end for

B. Outlier Elimination

To tackle the outlier problem, our approach is to eventually define an *acceptable RSSI range*, on a per-landmark basis, for each AP associated with that landmark. Once such a range is defined, we can then eliminate outlier MNs by first determining their predicted location (landmark), using the conventional RSSI-nearest neighbor (**NN**) approach, and then checking if the actual RSSI value lies within this landmark's acceptable range.

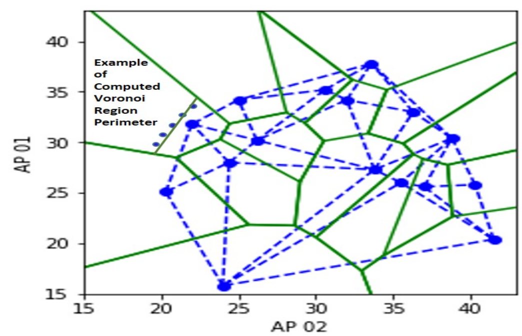

Fig. 4. Voronoi Tessellation for Residential Building

To create a boundary, we first divide each of the landmarks using an $N - dimensional$ (N= number of APs) Voronoi tessellation in the signal-space. Figure 4 shows such an example for $N = 2$ –i.e., with two APs. (In practice, for each landmark, we restrict ourselves to a 2-dimensional tessellation, involving the two APs with the strongest signal strength in the fingerprint DB. This was seen to provide sufficient practical discrimination and is computationally simple.) We also assume that we know the distance of representative *boundary points* of the fingerprinted region to each of the APs. Outlier elimination can consists of the following steps:

1) Boundary Estimation: The first step involves estimating the likely RSSI values at each of those boundary points. (Note that those points may not have been manually fingerprinted.) To estimate this, we utilize a path-loss propagation model:

$$P_{RSSI} = \beta - n.d + X_\alpha, \tag{3}$$

where P_{RSSI} is the RSSI strength, β is the transmitted power and antenna gains, n gives the path loss constant, d defines the distance and X_α is the shadow fading defined by the Gaussian random variable with zero mean. We first apply a regressor to the known AP-landmark distances and RSSI(landmark) readings to learn the optimal model parameters. Fig. 5 gives the straight line fit for the model for a given (landmark, AP) combination. The shadow fading coefficient is estimated by the average prediction error, according to $X_\alpha = \frac{\sum_{i=1}^{m}(P_{e_i} - P_{o_i})^2}{m}$, where P_{e_i} denotes the predicted value, P_{o_i} the observed value and m is the number of observed values.

Subsequently, we use the regressor to predict (without any additional fingerprinting) the RSSI readings at the representative boundary points, and then use the *average* of these values to denote the global minimum **per-AP** signal strength (α^G_{min})

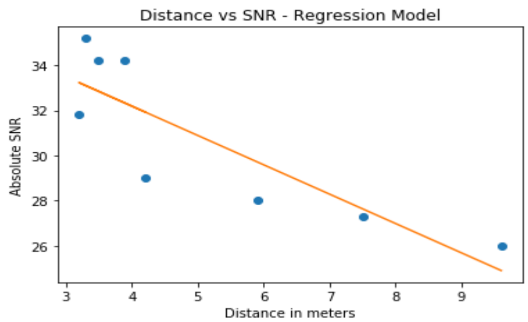

Fig. 5. SNR vs Distance Relation for Residential Building

that an MN located *anywhere* within the fingerprinting region should have. Similarly, we assume a minimum distance d_{min} (computed as the height of the floor, such that any legitimate point should be $> d_{min}$ away from the AP) and compute the global per-AP maximum permissible RSSI value α^G_{max} (using the regressor).

2) Legitimate Range Estimation: We next address the question: how do we define the real Voronoi region (in the signal space) for each landmark? To tackle this question, we need to define a range of legitimate RSSI values associated with each **(landmark, AP)** combination. In particular, the legitimate RSSI values, for AP_i, for an MN that has been mapped to landmark l are those that both (a) lie within the global range $(\alpha^G_{min}, \alpha^G_{max})$, *AND* (b) lie within the Voronoi region of landmark l. We compute the set of points (in the RSSI space) that satisfy both these criteria and accordingly define an additional, per-landmark set of thresholds for each AP: $\{\alpha^l_{min}, \alpha^l_{max}\}$. At the implementation level, we employ the Bowyer-Watson algorithm [22]. We feed the fingerprinted RSSI values for both the APs as input, which is subsequently used for creating the delaunay triangles. Further, we calculate the circumcenter by iterating over the three neighbouring triangle points.

C. Final Outlier Logic

Given the resulting Voronoi regions and a predicted landmark l for a test client, we determine its location estimate to be *legitimate* only if (i) if its RSSI readings satisfy the global thresholds $(\alpha^G_{min}, \alpha^G_{max})$ for every reporting AP, and (ii) if the RSSI readings are within the permitted Voronoi space of the landmark's two strongest APs (denoted by $AP1(l), AP2(l)$)– i.e., it satisfies the constraints:

$$\alpha^l_{min}(AP1(l)) \leq P_{RSSI_{AP1}} \leq \alpha^l_{max}(AP1(l)) \;\; \& \\ \alpha^l_{min}(AP2(l)) \leq P_{RSSI_{AP2}} \leq \alpha^l_{max}(AP2(l)). \quad (4)$$

Note that if a location is declared to be illegitimate, the client is then mapped to the subsequent (second-most likely) location; this process continues iteratively until a suitable and legitimate candidate location is found or until all landmarks are exhausted (in which case the MN's location is declared as *indeterminate*).

VI. Results and Discussion

In this section, we quantify the real-world performance of our improved AP cardinality and outlier detection methods, and show how they help improve the accuracy of occupancy detection.

A. Improvement in Cardinality

We first compare our modified algorithm (which includes so-called *stale* readings during the stationary periods of an MN) against the baseline approach, which ignore such stale readings. Fig 6 plots the location estimation error (the distance of the predicted location from the MN's ground truth) for 4 different clients (C1-C4), placed at different locations with the university research lab. We see that the appropriately curated inclusion of stale RSSI data helps to reduce the estimation error significantly, to less than 4 meters in at least 75% of these cases. The smaller improvement for C2 was due to the observed movement of multiple visitors through the area during the study, which affected the underlying radio environment.

Dominance & Accuracy of Stay Episodes Our proposed method is especially effective in tackling the cardinality problem during periods when the MN is stationary. To quantify the importance of improved localization during such stationary episodes, we analyzed the motion traces of all clients, in the Research Lab, over an entire day. We found that clients spent, on average, **92%** of the day in a "stationary state", with a mean stay duration of 372 minutes (and std. deviation. of 180 minutes). Moreover, to evaluate the possibility of our algorithm making false "stationary" inference, we experimented with multiple client devices that were (a) either completely stationary, or (b) made small movements (within 1-2 landmarks), over a 60 minute duration. We noted that the clients were classified as "stationary" (for the purposes of Algorithm 1) in 100% of all such cases: while small movements resulted in changes in the RSSI value reported by the associated AP, they do not actually cause MNs to generate explicit PROBE_REQUESTs.

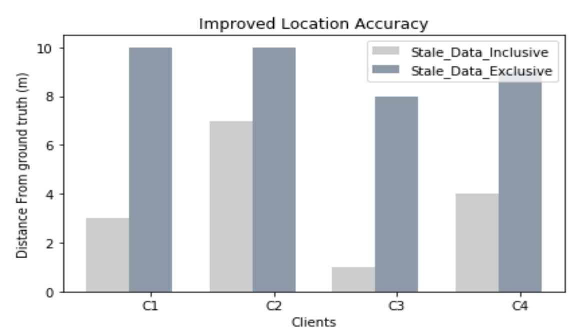

Fig. 6. Location Accuracy (with & without 'Stale' Readings)

To further demonstrate the benefit of improved cardinality, Table IV lists the location estimation accuracy vs. the cardinality of the corresponding RSSI reports (for the Research Lab). A location estimate is deemed to be accurate if the estimated location (landmark) is identical to the ground truth (the landmark nearest the MN's actual location.) Clearly, the augmentation of cardinality provides dramatic benefits, increasing the localization accuracy to ∼80-90% (in contrast to accuracy values less than 20% when cardinality=1).

TABLE IV
LOCALIZATION ACCURACY VS. CARDINALITY

Cardinality	**1**	**2**	**3**
Accuracy	15%	85%	92%

Our logic to improve the cardinality, and thereby the localization accuracy, normally distinguishes between stationary and moving devices by using RSSI reports from a 'new' AP. One can thus wonder if the approach is valid for scenarios where the entire region is covered by all the APs. To test this situation, we moved a test MN (shifting its location by 8m) around a region where all APs could be heard simultaneously. From Table V, we observe that at time (T) the AP01 and AP02 reported stale values. However, after the device was moved (at time=T+12), *both* APs generate fresh RSSI reports, with updated age values. Consequently, their stale entries are simply ignored and the device's estimated location is refreshed.

TABLE V
CLIENT MOBILITY WITHIN THE AREAS COVERED BY ALL APS

Timestamp	**Age**	**AP MAC**	**RSSI**	**Associated Status**	**Channel**
T	33	AP01	-50	0	153
T+2	20	AP02	-65	1	40
T+12	6	AP01	-57	0	153
T+14	1	AP02	-67	1	40

B. Location Outlier Elimination

As explained earlier, the outlier algorithm discards those location estimates that lie outside a permitted signal strength range defined for each (AP, landmark) combination. As an illustration of the effectiveness of this strategy, Table VI lists the SNR values for a particular client that was reported at a landmark in the residential building, even though it was *actually placed at a point outside the fingerprinted region*. The table lists the SNR value for the different APs. As observed, the client readings are eliminated in all three cases, as the SNR values are seen to be below the AP-specific *global* thresholds.

TABLE VI
DATA REPORTED FOR TEST CLIENT IN THE RESIDENTIAL BUILDING

Access Points	**Client SNR**	**Without Threshold**	**With Threshold**
AP01	11	Included	$\alpha^G_{min}(1) = 13.5$; Excluded
AP02	17	Included	$\alpha^G_{min}(2) = 20.3$; Excluded
AP03	15	Included	$\alpha^G_{min}(3) = 19.5$; Excluded

Robustness of Outlier Detection: To test the ability of our technique to accurately separate outliers from legitimate estimates, we conducted studies where an MN was placed at multiple distinct locations under *varying crowdedness levels* (both (a) close to the boundary, but *within* the fingerprinted Research Lab, and (b) *outside* but near the Lab). Table VII plots the outlier detection results under both conditions, and shows that our outlier detection is robust (overall accuracy=84%), with both low false-positive and false-negative rates.

TABLE VII
OUTLIER DETECTOR PERFORMANCE

Ground Truth	**Inferred-Legitimate**	**Inferred-Outlier**
Inside Boundary	98%	2%
Outside Boundary	14%	86%

C. Occupancy Estimation

Direct, large-scale validation of the outlier detector logic is difficult, due to the difficulty in obtaining ground truth. Instead, we now provide an indirect evidence of the benefit of such outlier elimination–namely, a dramatic improvement in the estimation of *overall occupancy*.

To study the occupancy estimation outcome, we manually recorded the total occupancy in the public area of the residential building over a 1-hour observation window. We further differentiated the occupancy estimates across 3 different landmarks: L_8, L_{16} and L_{17}. Figure 7 plots the ground truth as well as the estimate occupancy values, both with and without our threshold-based outlier elimination mechanism. Consistent with the prior exemplar (Table III), we see that, without the threshold, the occupancy is consistently over-estimated, as the count includes several devices that attach to the indoor APs, even though the users are located outside the building. We observed that the introduction of our outlier-based elimination process results in an at least 80% reduction in the occupancy estimation error, across all 3 landmarks.

On closer examination, we see that the improvement in estimation accuracy is more dramatic for landmarks L_{16} and L_{17}. These two locations lie on the boundary of the building, making it more likely for them to include extraneous MNs that connect to the APs but are actually located outside. In contrast, for landmark $L8$, we see that, in the absence of the threshold-based filtering, the occupancy count (ground truth=1) is erroneously estimated to be 0. This implies that, the device actually located at L_8 was incorrectly mapped to a different landmark. This example illustrates a secondary benefit of outlier detection: it not only eliminates devices located outside the fingerprinted area, but also helps to improve the accuracy of location estimation by rejecting location estimates that violate *landmark-specific thresholds* (Equation 4). An analysis of a 9-hour data trace showed that we were able to eliminate (as outliers) approx. 18% of the clients that were initially mapped to a location within the fingerprinted areas of the building.

VII. CONCLUSION

In this paper we tackled two practical problems existing in the implementation of accurate server-side indoor localization

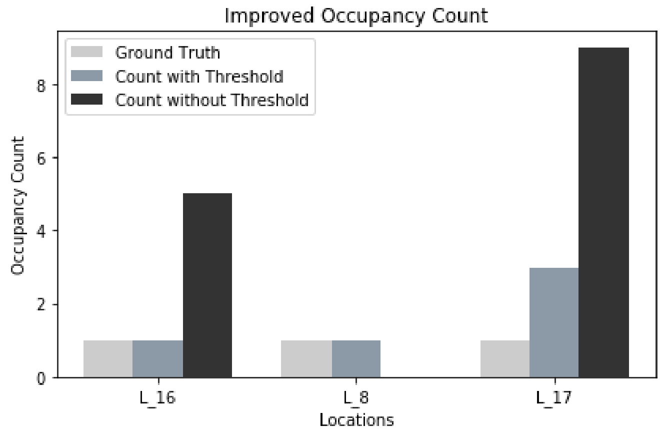

Fig. 7. Occupancy Count - With/Without Filtering Thresholds

(using both Aruba and Cisco equipment): (a) low AP cardinality for stationary clients and (b) over-counting of devices located outside the fingerprinted area. While the first problem might be less acute in venues where users typically have low residency times (e.g., in train stations), the second problem is universal. We tackled the low cardinality problem by explicitly delineating stationary periods for each MN, and using hitherto-discarded *stale* reports from other APs to augment the cardinality of AP measurements. We also tackled the over-counting or mis-attribution problem by effectively using a Voronoi-tessellation approach to devise a *per-landmark, per-AP* acceptable range of signal strength readings, and iteratively discard location estimates that did not conform to these ranges. Our empirical results prove that these mechanisms have clear and significant practical impact: (a) they reduce the average localization error for stationary clients from ∼8-10meter to 2-4meter (such stationary behavior is observed 92% of the time); and (b) reduce the error of aggregated occupancy estimates by 80+%. In ongoing work, we are integrating these mechanisms into our occupancy monitoring system, and shall then conduct significantly larger, longer-scale studies to quantify the overall robustness of the proposed techniques.

Acknowledgement

This work is supported by the National Research Foundation, Prime Ministers Office, Singapore under its International Research Centres in Singapore Funding Initiative. The authors gratefully acknowledge the support of researchers from SMU's erstwhile LiveLabs research center, who helped provide the base WiFi RSSI data.

References

[1] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless rssi fingerprinting localization," *Signal Processing*, vol. 131, pp. 235 – 244, 2017.

[2] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the mona lisa: Spot localization using phy layer information," in *10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12, 2012.

[3] S. Sen, D. Kim, S. Laroche, K.-H. Kim, and J. Lee, "Bringing cupid indoor positioning system to practice," in *Proceedings of the 24th International Conference on World Wide Web*, ser. WWW '15, 2015.

[4] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal, "Sentinel: Occupancy based hvac actuation using existing wifi infrastructure within commercial buildings," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '13, 2013.

[5] B. S. Ciftler, S. Dikmese, . Gven, K. Akkaya, and A. Kadri, "Occupancy counting with burst and intermittent signals in smart buildings," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 724–735, April 2018.

[6] R. K. Balan, A. Misra, and Y. Lee, "Livelabs: Building an in-situ real-time mobile experimentation testbed," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, ser. HotMobile '14, 2014.

[7] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, vol. 2, March 2000, pp. 775–784 vol.2.

[8] D. Jaisinghani, R. K. Balan, V. Naik, A. Misra, and Y. Lee, "Experiences & challenges with server-side wifi indoor localization using existing infrastructure," in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MobiQuitous '18. New York, NY, USA: ACM, 2018, pp. 226–235.

[9] M. Youssef and A. Agrawala, "The horus location determination system," *Wireless Networks*, vol. 14, no. 3, pp. 357–374, Jun. 2008.

[10] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *10th USENIX Conference on Networked Systems Design and Implementation*, ser. nsdi'13, 2013.

[11] B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless lan," in *LAN, First IEEE International Conference on Wireless broadband and Ultra Wideband Communications*, 2007, pp. 13–16.

[12] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.*, Aug 2004, pp. 14–23.

[13] S. Mazuelas, A. Bahillo, R. M. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. J. Abril, "Robust indoor positioning provided by real-time rssi values in unmodified wlan networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 821–831, Oct 2009.

[14] J. Talvitie, M. Renfors, and E. S. Lohan, "Distance-based interpolation and extrapolation methods for rss-based localization with indoor wireless signals," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1340–1353, April 2015.

[15] K. El-Kafrawy, M. Youssef, A. El-Keyi, and A. Naguib, "Propagation modeling for accurate indoor wlan rss-based localization," in *2010 IEEE 72nd Vehicular Technology Conference - Fall*, Sep. 2010, pp. 1–5.

[16] M. B. Kjaergaard, "Indoor location fingerprinting with heterogeneous clients," *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31 – 43, 2011.

[17] I. Meraki, "Location analytics." [Online]. Available: https://meraki.cisco.com/solutions/location-analytics

[18] A. Networks, "Rtls - integrating with the rtls data feed," pp. 1–14, 2012. [Online]. Available: https://community.arubanetworks.com/aruba/attachments/aruba/unified-wired-wireless-access/23715/1/RTLS_integrationv6.docx

[19] X. Hu, L. Song, D. Van Bruggen, and A. Striegel, "Is there wifi yet?: How aggressive probe requests deteriorate energy and throughput," in *Proceedings of the 2015 Internet Measurement Conference*, ser. IMC '15. New York, NY, USA: ACM, 2015, pp. 317–323.

[20] J. Freudiger, "How talkative is your mobile device?: An experimental study of wi-fi probe requests," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec '15. New York, NY, USA: ACM, 2015, pp. 8:1–8:6.

[21] D. Jaisinghani, V. S. Naik, S. K. Kaul, R. Balan, and S. Roy, "Improving the performance of wlans by reducing unnecessary active scans," *CoRR*, vol. abs/1807.05523, 2018.

[22] D. F. Watson, "Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes*," *The Computer Journal*, vol. 24, no. 2, pp. 167–172, 01 1981. [Online]. Available: https://doi.org/10.1093/comjnl/24.2.167