Do WiFi Probe-Requests Reveal Your Trajectory?

Abhishek Kumar Mishra, Aline Carneiro Viana, Nadjib Achir

Do WiFi Probe-Requests Reveal Your Trajectory?

Abhishek Kumar Mishra*, Aline Carneiro Viana*, Nadjib Achir*†
* *Inria*, France † *University Sorbonne Paris Nord*, France

***Abstract*—In this paper, we propose the first framework that introduces the concept of the user's bounded trajectory. We propose to leverage the signal strength of users' public WiFi probe requests collected from measurements of multiple deployed WiFi sniffers. First, we investigate and characterize errors in RSSI-based radial-distance (between the user and each sniffer) estimation. Then, we approximate such radial distances leverage and deduce bounds associated with a user's position. Finally, we infer a user's bounded trajectory using the spatiotemporal bounds of users' locations over time. We guarantee the bounds to enclose a user in space and time, with 95% confidence and a 10% margin of error. Using real-world and large-scale synthetic datasets under heterogeneous contexts and wireless conditions, we infer trajectories with bounds' *width* of less than $10m$ in 70% of cases with users' *inclusiveness* close to 100%.**

***Index Terms*—Passive sniffing, RSSI, distance estimation**

I. Introduction

User-Trajectory investigation has attracted considerable literature attention. With mobile phones becoming proxies for human presence, network resources have been exploited to investigate users' mobility [1]. Spatio-temporal mobility datasets are nowadays acknowledged as a common tool to study users' trajectories: e.g., mobile phone records [2], WiFi [3], [4], and BLE data [5]. Most require access to users' devices, network infrastructure, or volunteer recruiting, while some lack scalability. Instead, in this paper, *we base our investigation on datasets that are non-intrusive and independent of third parties*. *We focus on WiFi passive sniffing*.

RSSI (Received Signal Strength Indication) is the most widely used localization metric, with no hardware-specific requirements. In the indoor controlled environment, previous works attempt to localize users through RSSI-based distance measurements [6], [7]. On the other side, in outdoor scenarios, there has been an attempt to use machine learning techniques to lower the distance-estimation error [8]. But errors still go up 16m, that too in semi-controlled environments. The literature lacks works leveraging RSSI for outdoor users' localization, mainly due to its imposed uncertainties. Uncertainties incur significant errors in distance estimation, making user localization a challenging task.

Solutions using RSSI for estimating users' trajectories contain localization errors arising from inaccurately measured distances between the user and the regarded sniffer. Past works have tried to mitigate such errors to some extent using smoothing techniques for RSSI [9]. However, smoothing itself causes bias in the estimated distance. Another approach [10] adapts the path-loss model parameters by deploying fixed reference nodes and reference tags placed at known positions. This approach requires extra hardware and is primarily developed for indoor localization. Although the employed effort, errors in estimation remain an issue. *To the best of our knowledge, none of the solutions focus on the characterization of localization errors in outdoor scenarios.*

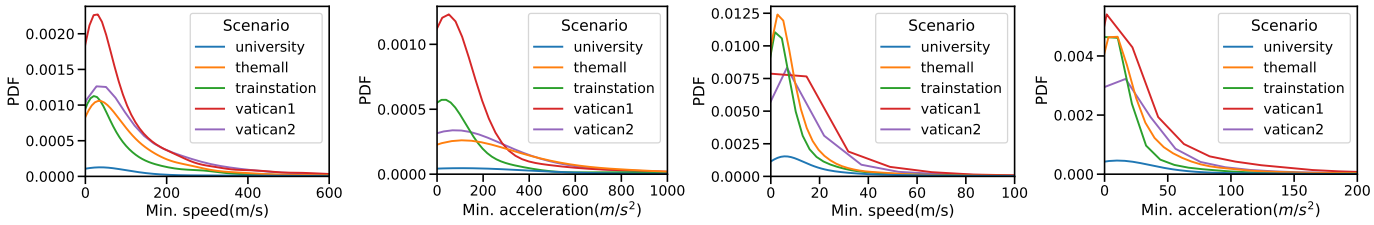
Our approach relies on an in-depth investigation and novel modeling of *(i) the severity, (ii) the non-stationary behaviour, and, (iii) the variability of RSSI-based distance-estimation errors* in *outdoor* passive measurements (cf. Sec. II). The resulting error's comprehension leads us to identify and formalize *a span of possible user locations* associated with measurements in different time intervals. We denominate such span of possibilities as *bounds of locations* of a user (cf. Sec.s III and IV). Bounded locations are then used to infer bounds of users' full-fledged trajectories over time that we designate as the user's *bounded trajectory* (cf. Sec. V). This work does not raise any ethical issues.

II. Challenges in exploiting RSSI

To investigate the credibility of raw RSSI, we look at the measured RSSI of collected probe requests. To figure out its potential for 2D spatial estimation of users' location, we build upon literature solutions to the *radial-distance* estimation (i.e., distance in meters separating users from sniffers) from packets' RSSI. We use the public anonymized *Sapienza* trace [11] , describing a large number of WiFi probe requests passively collected in eight different contextual scenarios. We choose datasets from five scenarios: *vatican 1*, *vatican 2*, *university*, *trainstation*, and, *themall*. The three first scenarios denote dense outdoor environments, while the last two showcases highly frequented, ample public spaces.

Using the raw RSSI values in the five selected datasets, we estimate radial distances through the log-distance path-loss model configured with standard parameters. As we do not know the actual "severity" of the physical environments where probes were collected, we consider a moderate path-loss exponent (γ=3.0). For *Sapienza* datasets, no *ground-truth* with true devices-to-sniffers radial-distances is available. Obtaining this *ground-truth* is challenging, with large-scale passively collected traces with mobile users. Hence, to verify the correctness of the computed radial distances, we first obtain the corresponding range values of (i) instantaneous *speeds* – indicating users' changing behaviors in estimated radial distances over time – and (ii) *accelerations* –, indicating the geographical dynamism of users mobility, per time window

This work has been partially funded by the ANR MITIK project, French National Research Agency (ANR), PRC AAPG2019.



(a) (b) (c) (d)

Fig. 1: min. speed (a), (c) and min. acceleration. (b), (d) from raw and smoothed RSSI values.

of 50ms. This results in minimum values of the actual user's speed and acceleration, which we compare with the corresponding realistic values of human mobility.

Fig. 1a reveals the range of minimum speeds issued from the radial distances estimated from datasets' raw RSSI values, considering path-loss severity of $\gamma = 3.0$. Similarly, Fig. 1b shows the range of minimum accelerations. We note that the minimum speed and acceleration values reach around $200m/s$ and $800m/s^2$. Those are unrealistic speed and acceleration values, indicating errors in the radial-distance estimations associated with datasets' raw RSSI values.

To reduce the impact of high RSSI fluctuations, filters have been suggested in the literature to smooth noisy raw RSSI values that cause distance-estimation errors [1], [12]. Smoothing approaches include feedback, moving-average, or median filters, among others. We choose the Exponentially Weighted Moving Average (EWMA) filter with a smoothing factor of $\alpha = 0.1$, which gives exponentially decreasing weights for observed raw RSSI values over time. EWMA is considered the state-of-the-art model to capture losses in densely populated areas.

Fig. 1c reveals the range of minimum speeds, and Fig. 1d shows the corresponding range of minimum accelerations when applying smoothing. We notice that the minimum smoothed speed and corresponding acceleration reduce and reach $60m/s$ and $150m/s^2$. Although still not corresponding to the range of realistic pedestrian speeds and accelerations.

In conclusion, we observe that even smoothing incurs unrealistic induced minimum speed and acceleration values, illustrating that the radial-distance estimation is incorrect. The incorrect estimation is attributed to the noisy RSSI observations in dense outdoor scenarios. It highlights *the current difficulties in using RSSI for accurate estimation of users locations and consequently, trajectories.*

III. The Proposed Framework

In this paper, we *redefine the problem of trajectory estimation of users by considering the "bounds" on the range of locations where users are expected to be present at a certain point of time.* The goal is to deduce a user's *bounds of locations* that is probabilistically guaranteed (confidence) for it to be located within a certain distance from the actual locations visited by the user, i.e., the *ground truth*.

A. The framework

The first component of the framework generates **Observation Sets**. We divide the duration for which sniffers "observe" a user into time intervals, noted as δt. Each time interval describes an *observation set* of sniffers with respect to a user. We carefully settle the value of δt to fulfill the following conditions. On the one hand, for a more precise 2-dimension localization and a good representation of variability of incurred environment errors, δt has to be large enough to (i) enable multiple sniffers to capture a user's presence and (ii) allow observing enough probe request samples in its time interval. On the other hand, to refrain from considerable changes in the user position, the value of δt must be limited. We satisfy this compromise by empirically fixing the value of δt to 4s.

The second component performs the **Error Characterization** (Sec. IV), i.e., the characterization of radial-distance errors from observation sets. For the first time in literature, we identify that estimation errors can be viewed as a sum of two additive error components. The first component denotes the mean error behavior incurred in radial distances estimations, which we call *span error*. The second component captures the fluctuations in distance estimations originating from the changing RSSI values of a user in short time intervals, which we call *environment error*. We quantify these fluctuations and estimate their distribution in each observation set.

The last component of the framework generates **Bounded Trajectories** (Sec. V). First, we do a Gaussian fit to *environment error* distributions. Then, to obtain the total estimation error, we sample multiple times from the resulting fitted distribution and add the error samples to the corresponding *span error*. Next, we add the obtained total errors to users' radial-distances and feed them to our multilateration-based location estimator. Depending on the number of samples drawn from the fitted environment error, we obtain a set of user positions named *bounds of locations*. Finally, we aggregate them per observation set to get *bounded trajectories* of users.

B. The datasets

We use real fully anonymized datasets for calibration purposes, collected nearby the main entries of a large university campus with a high number of passers-by, named *Campus 1* and *Campus 2*. At the collection procedure, a controlled WiFi probe device and one nearby passive sniffer were deployed at

fixed and known positions. Finally, the released *Campus 1* and *Campus 2* datasets only describe timestamped probe-request packets only from the controlled probe device.

We introduce *WiSurve*, a emulator which creates real-world, large-scale passive sniffing environments for WiFi with *ground truth*. *WiSurve* takes real geographical areas as input road-network using OpenStreetMap[1]. It generates pedestrian trajectories with varied individual characteristics on top of the input road network using SUMO[2]. Finally, it enhances the state-of-the-art WiFi simulator *NS-3*[3] to mimic WiFi standard networks composed of a set of sniffers, access points, and mobile nodes.

IV. Error Characterization

A. Converting RSSI to radial-distances

We use the log-distance path loss model (Eq. 1) to capture losses encountered for signals in densely populated areas. In Eq. 1, PL is the total path loss in decibels (dB) at a distance r_{ji} between the user j and the sniffer i. $P_{Tx_{dBm}}$ and $P_{Rx_{dBm}}$ are the transmitted and received power in dBm. PL_0 is the path-loss (dB) at reference distance, r_0. γ is the path loss exponent that depends on the propagation characteristics of the received signal. Finally, X_g is a normal random variable that considers losses (e.g., obstructions caused by buildings or pedestrians) from shadow-fading in outdoor scenarios.

$$PL = P_{Tx_{dBm}} - P_{Rx_{dBm}} = PL_0 + 10\gamma \log \frac{r_{ji}}{r_0} + \mathcal{X}_g \quad (1)$$

We set the reference distance r_0 to be 1 m. Using Eq. 1, we compute the approximate radial distances (r_{ji}) between users and sniffers by using RSSI values directly observed from real WiFi probe requests, as in Eq. 2.

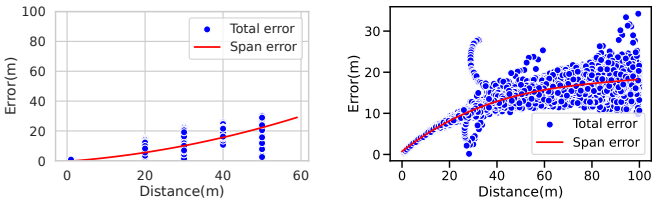
$$r_{ji} = 10^{\frac{P_{Tx_{dBm}} - P_{Rx_{dBm}} - PL_0 - \mathcal{X}_g}{10\gamma}} \quad (2)$$

The computation of the radial distance r_{ji} is conditioned to the estimation of thenoptimal values of γ and $\mathcal{X}_g$ for the considered wireless environment, as in Eq. 2. The other inputs to the equation are either standard: $P_{Tx_{dBm}}$ (set to 23 dBm) and PL_0 (set to 46.67 dB), or known: $P_{Rx_{dBm}}$ (observed RSSI). Values of γ vary from around 2.1 in a line-of-sight (LoS) environment to close to 4 in dense scenarios. Hence, we set the range of γ values, i.e., R_γ, to be $[2.1, 4.0]$, corresponding to the range of values of path-loss exponents considered in optimal parameter search.

To consider shadow fading, we model $\mathcal{X}_g$ as the random variable with a Gaussian distribution with zero mean (μ) and standard deviation (σ) in decibels. The σ varies from close to 0 in free space to around 5 for signals inside commercial stores. We set the search-space for the standard deviation of the random variable, i.e., $R_{\mathcal{X}_g}$, to be $[0, 5.0]$.

To obtain γ_{opt} and $\mathcal{X}_{g_{opt}}$, i.e., optimal path-loss parameters for a particular sniffing environment, we introduce a calibration step that reduces the differences between actual and modeled values. Our calibration methodology uses the real *Campus 1* and *Campus 2* datasets. The obtained values from the calibration give the optimal path-loss parameters for simulated environments (*WiSurve dataset*).

Basically, we generate a number of uniformly-spaced samples (i.e., S_γ and $S_{\mathcal{X}_g}$) from the ranges, R_γ and $R_{\mathcal{X}_g}$. For each combination (i.e., $S_{\gamma,\mathcal{X}_g}$) of drawn samples, we calculate the *distance-estimation error*. We define the error as the average of the differences between the distances calculated from the RSSI values (see Eq. 2) and the corresponding ground-truth distances (cf. Sec. III-B). Finally, we obtain optimal parameters ($\gamma_{opt}, \mathcal{X}_{g_{opt}}$) by finding the parameter combination for which the resulting distance-estimation error is the lowest.



(a) *Campus 1* (b) *WiSurve dataset*

Fig. 2: Errors in radial-distance estimation.

B. Formalising distance-estimation errors

Based on our previous investigations on errors' causes and calibration observations, we form two hypotheses:

Hypothesis 1: The average estimation error increases with an increase in the radial distance.

Hypothesis 2: The error fluctuations in an observation set (δt) grow with increasing radial distances. Their distribution can be modeled in small intervals of time.

To attest to those two hypotheses, we extensively investigate, through two case studies, the behavior of radial-distance errors when approximated with the log-distance path-loss model having optimum parameters (cf. Sec. IV-A).

1st case study: We infer radial-distance errors from the RSSI values described in *Campus dataset 1* and *Campus dataset 2*. Their controlled scenarios allow us to evaluate errors' behaviors when considering fixed and discrete distances between a probe source and a sniffer.

2nd case study: We exploit a synthetically generated *WiSurve dataset*, introducing a mobile scenario with varying distances between multiple users and sniffers. Here, we have 100 users that move with a walking speed of $1m/s$ in the sniffing zone while emitting probe-requests.

Results from Fig. 2a and Fig. 2b obtained for the two case studies show error behaviors in the distance estimation and related fluctuations as a function of source-to-sniffer radial distances. Related to Hypothesis 1, we observe a consistent trend followed by the average error with respect to the radial-distance. As stated in Hypothesis 1, estimation error grows with increasing radial-distances. We further show that this trend can be predicted through a polynomial approximation. Related to the Hypothesis 2, results also show that distance-estimation errors' fluctuations increase with larger distances.

[1] www.openstreetmap.org [2] https://sumo.dlr.de [3] https://www.nsnam.org/

Finally, we also observe that their magnitude also changes according to the sniffing environment.

Hence, we decompose distance estimation errors into *span error* and *environment error*. To obtain the span error, we calculate the error for each RSSI trace element as the difference between the distance obtained from RSSI using the optimal path-loss model and the corresponding ground-truth distances. Finally, using a polynomial fit, we estimate the average error, i.e., the *span error*, at a given radial-distance from the user. Using *Campus dataset 1*, *Campus dataset 2*, and *WiSurve dataset*, we find that polynomials of order 3 are the best fit to characterize *span errors*. Higher orders fail to catch the rate of error change with distance.

When considering the total error in the estimated source-to-sniffer distances, *environment error* is seen as a *"complement"* to the span error in the corresponding observation set. It varies with each observation set in the user-trajectory. Hence, we define environment error for each user, across all observation sets, per sniffer. Due to sniffers' different spatial positioning, we observe varied environment fluctuations in an observation set. To estimate the environment error, we discretize the per-user trace in time to get the *observation sets*. Then, per the observation set, we calculate the user-to-sniffer distances and subtract the corresponding mean span error per-sniffer. This "residue" captures the fluctuation of estimation errors in an observation set, which we name as *environment error*.

V. Bounded trajectories

A. Formalizing the localization error

In the following, we use *multi-lateration* for the positioning for user positioning [13]. Multi-lateration uses measured distances between users and multiple sniffers at known locations.

Let's consider N_s the number of sniffers receiving probe-requests of a user j during the k'th observation set. We define the corresponding estimated position of a user j as $T_j = (x_j, y_j)$. We represent the location of a sniffer i in two-dimensions as $S_i = (x_i, y_i)$, and the radial distance between user j and sniffer i in k'th observation set as R^k_{ji}:

$$R^k_{ji} = d(T_j, S_i) = \sqrt{\sum_{m=1}^{2} \left(T^m_j - S^m_i\right)^2} \tag{3}$$

We also define the measured distance from the probe-request RSSI's between user j and sniffer i in the observer set as r^k_{ji}. Finally, we state the localization error as $LocErr^k_j$:

$$LocErr^k_j = \sqrt{\sum_{i=1}^{N_s} \left(R^k_{ji} - r^k_{ji}\right)^2} \tag{4}$$

To find the optimal user-location, T^*_j, of user j in an observation set k that minimizes the localization error, we need to solve a non-linear least squares problem for the minimization of $LocErr^k_j$. We formulate that problem as:

$$T^*_j = \arg\min_{T_j} LocErr^k_j \tag{5}$$

For Eq. 5, one option is a closed-form solution, which generally can not be considered when circles in multi-lateration do not intersect at one common point. Numerical methods are a strong candidate for getting the *best approximate* solution, even in the condition of not-intersecting circles. We use simulated annealing (SA) to obtain the global minimum with sufficient iterations.

B. Finding bounded trajectories

To get the bounds of users' locations, we include estimated radial-distance errors while finding optimal user-locations.

Modelling environment errors' distribution: Constraints on the length of δt may lead to an insufficient number of probe requests for characterizing the environment error distribution (cf. Sec. III-A). To tackle the probe sample's scarcity, we estimate the actual distribution of environment error distribution ($\mu = 0, \sigma = \sigma_{N\_distr}$) through the best Normal fit. We empirically find that for the δt with high number of samples, the Normal fit is indeed close to the actual distribution. The estimated distribution allows us to sample and incorporate localization errors in δt. We sample a $Depth$ number of times to identify the bounds on the possible user locations.

Bounds' probabilistic guarantee: We define $Depth$ such that the resulting bounds have a level of confidence, **Z** and a *margin of error*, MOE [14]:

$$Depth = \frac{\mathbf{Z}^2 \sigma^2_{N\_distr}}{MOE^2} \tag{6}$$

where σ_{N_distr} varies with every δt. We set the level of confidence to 95 % and the margin of error to 10%.

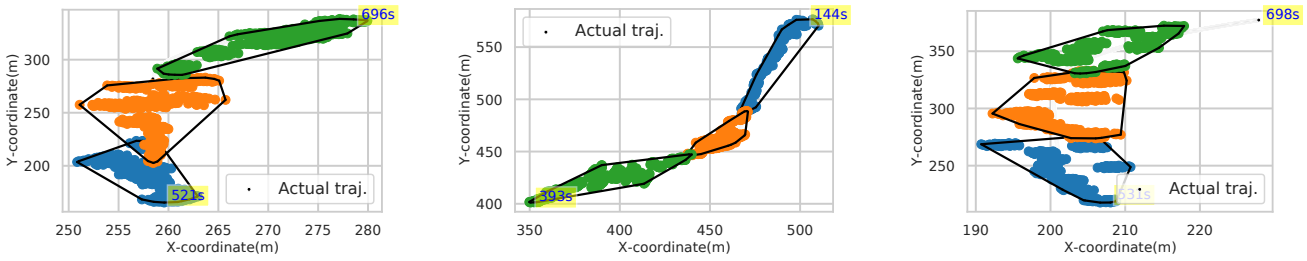
Getting bounds of users' locations: We first obtain span and environment errors per sniffer. The sum of the mean span error and the environment errors sample drawn from the fitted Normal distribution results in the total error. Next, we get the possible user-distances, by adding the total error to approximate radial-distances. Then, we use Simulated Annealing (SA) to obtain the optimal target location (2-D) of the user from the possible user distances. Finally, we repeat the finding of optimal user locations a $Depth$ number of times to obtain a set of points, for which the enclosure gives us the *bounds* of user locations in the considered observation set (δt).

Bounded trajectory: Enclosing the above-obtained set of optimal-target points (i.e., $Bounds$), we define the bounded trajectory $B_{user_j} = \{\mathcal{C}_{j1}, \mathcal{C}_{j2}, \mathcal{C}_{j3}, ..., \mathcal{C}_{jk}, ..., \mathcal{C}_{jM}\}$, as a sequence of convex hulls ordered in time, spanning the trajectory. $\mathcal{C}_{jk}$ represents the kth member of the sequence of convex hulls when considering the trajectory of the jth user. We divide the user trajectory into multiple (three) convex hulls to capture the turns (changes in direction) that a user takes over time.

VI. Evaluation

A. WiSurve datasets

Scenario specificity: The nature of bounded trajectories depends upon the density of captured probe requests, which relies on the number of deployed sniffers. Another factor is the average number of missed packets, which may increase



(a) With default metrics. (b) With high users' number. (c) With low sniffers' number.

Fig. 3: Bounded trajectories' illustration.

with the increase in user density since sniffers are limited to listening to one channel at a time. Finally, the time a user spends in the sniffers' view is a function of her mobility behavior. We generate *WiSurve* datasets varying: the number of (i) sniffers, (ii) users, and (iii) user speeds.

WiSurve parameters: We consider a passive-sniffing Manhattan area of size $600m$ x $600m$ tessellated in grids of size $20m$ x $20m$. We generate trajectories with users performing random walk inside the tessellated area. We keep the grid size smaller than the WiFi range (i.e., we consider it $25m$) and deploy sniffers on randomly-chosen vertices. We set devices to use 802.11n (2.4 GHz) WiFi standard with a simulation duration of $40min$. We use a real-world data-based outdoor path-loss model [15] (cf. Sec. III-B). We vary user speeds from walking ($1m/s$) to running ($2.6m/s$). The *default* scenario consists of 80 sniffers and 150 walking users.

B. Illustrating bounded trajectories

Fig. 3a illustrates the bounded user-trajectories in the default scenario. We observe that convex hulls completely bound the actual trajectory of the user for its entire duration in the sniffing zone ($521s$ to $696s$). Fig. 3c certifies the successful identification of bounds of the user trajectory even for reduced number of sniffers (25), when having a higher number of missed packets. Although overshadowed by the points that convex hulls enclose, Fig. 3 shows that the actual (*ground-truth*) trajectories inside the sniffing zone are fully captured within the revealed bounds. We also verify our framework against an increased number of users (200) (Fig. 3b) and for higher speeds ($2.6m/s$). We do not illustrate all of them in this paper due to space constraints.

C. Quality of inferred bounds

We state the quality of bounded trajectory is expressed through two metrics (i) *correctness* and (ii) *width*. *Correctness* brings the intuition of "inclusiveness," where a user's bound of locations includes its real precise locations. The *width* captures how narrow the bounded trajectory is to limit the extent of possible user locations.

Bounds' correctness: We infer the "inclusiveness" of bounds as the length (in terms of the period of time) of the actual trajectory enclosed by the convex hulls ($\mathcal{C}_{jk}$) of a user j. Given time intervals for which a user was observed during its ideal trajectory, we define "ideal inclusiveness" as the length of the actual trajectory enclosed between the minimum and maximum timestamps of $\mathcal{C}_{jk}$. Note that the "ideal inclusiveness" acts as *ground-truth* for "inclusiveness". Finally, we compute the *Correctness* as the distance between the distributions of resulting bounds' "inclusiveness" and "ideal inclusiveness".

Fig. 4a, 4b, and 4c investigate the length of actual trajectories inside the bounded ones (per hull), which depict "inclusiveness". In Fig. 4a and 4b, the distributions of enclosed trajectory-lengths are practically the same irrespective to varying sniffers and users in the zone. This validates the resilience of our framework against sniffer and user densities. Fig. 4c demonstrates that the enclosed length of the actual trajectories decreases when increasing user speeds, as highly mobile users stay in the sniffers' range for shorter periods.

We find the Hellinger distance between distributions of *inclusiveness* and *ideal inclusiveness* for the quantitative analysis of the *correctness*. We get a distances below 0.22 for all the scenarios; showing that the corresponding distribution-pairs are very similar. This validates the *correctness* of our framework's bounded trajectories.

Bounds' width: To ensure the "utility" of the bounded trajectories, its resulting *width* has to be relatively narrow. We define the *width* of the bounded trajectory of a user j, as:

$$Width\left(B_{user_j}\right) = \frac{\sum_{k=1}^{M}\left(ActualTrajDist(\mathcal{C}_{jk})\right)}{M} \quad (7)$$

where $ActualTrajDist$ calculates the distance between each of the enclosed points of the hull to the corresponding closest-in-time point on the actual trajectory. It gives us a measure of how "spread" are the hulls' enclosed points, i.e., bounded trajectory around the actual trajectory. We average out the "spread" over all the hulls of a user j to obtain the *width*.

Fig. 4d, 4e, and 4f report the *width* of bounded trajectories, for scenarios with varying number of deployed sniffers, user-densities, and user speeds. Smaller is the number of sniffers, fewer is the perceived probe-requests in an observation set, and the more significant span error on average. Nevertheless, Fig. 4d shows the resilience of bounds' width to the variation of average total distance-estimation errors as bounds' widths are almost the same for more than 25 sniffers.

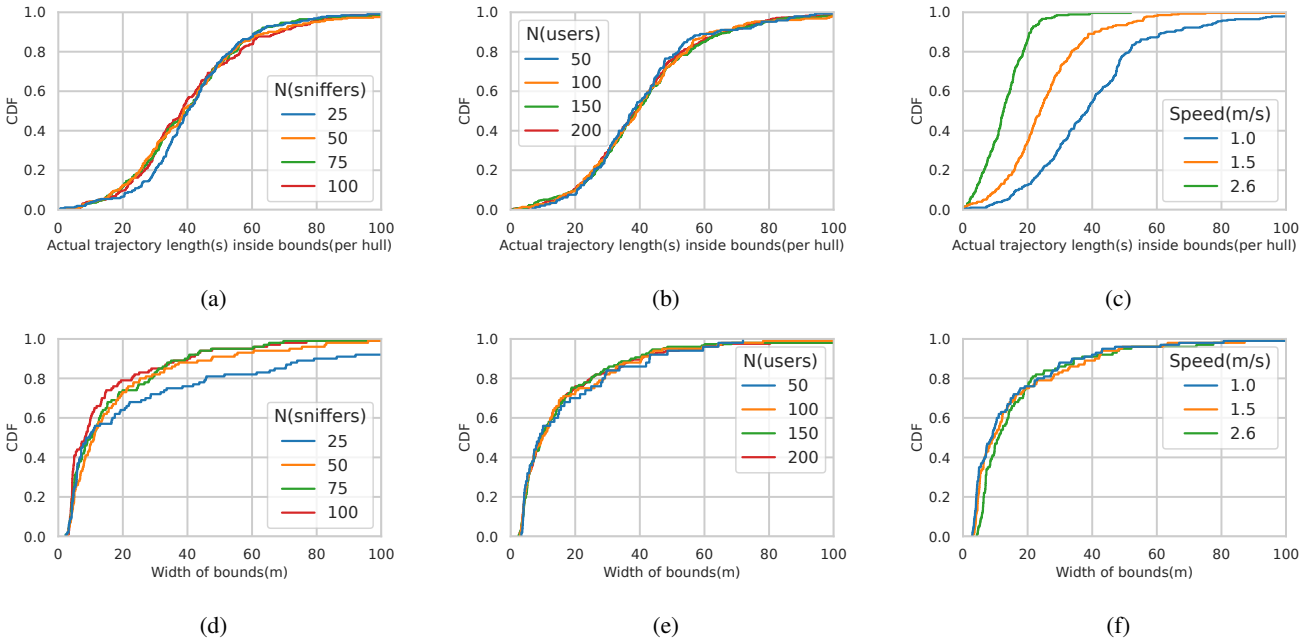


Fig. 4: (a), (b), (c): *Inclusiveness* per hull. (d), (e), (f): Bounds' width.

Increasing the number of users and speed increases the magnitude of localization errors due to the higher number of missed probe-requests and lower observation sets, respectively. Fig. 4e and 4f shows that, this increase in error amounts has a negligible decrease in the width. In essence, Fig. 4d demonstrates that the *width* of bounded trajectories is less than $10m$ for 70% of users, when considering enough sniffers (i.e., 50 or more). This *width* is 8.3% of the maximum WiFi range ($140m$) seen in utilized large-scale *WiSurve dataset*. These results certify the "utility" of bounded trajectories.

VII. Conclusion

Estimating trajectories of mobile users through non-intrusive measures remains an open problem. This paper leverages RSSI as a localisation metric and *transforms* related errors in distance-estimation into an "insightful metric". We show distance-estimation errors are a combination of two distinct components, i.e., span and environment errors. We then use this classification to find *correct* and *useful* bounds of a user's location in short time intervals, which are then exploited to reveal the bounds of the user's trajectory. This work does not raise any ethical issues.

References

[1] G. Deak, K. Curran, and J. Condell, "Evaluation of smoothing algorithms for a rssi-based device-free passive localisation," in *Image Processing and Communications Challenges 2*, Springer, 2010, pp. 469–476.

[2] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[3] M. Čavojskỳ, M. Uhlar, M. Ivanis, *et al.*, "User trajectory extraction based on wifi scanning," in *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, IEEE, 2018, pp. 115–120.

[4] A. Trivedi, C. Zakaria, R. Balan, *et al.*, "Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–26, 2021.

[5] G. Bernard, C. Faucher, and K. Bertet, "Towards reconstruction of human trajectories in indoor environments.," in *EKAW (Posters & Demos)*, 2018, pp. 37–40.

[6] R. M. Susanti, K. M. Adhinugraha, S. Alamri, *et al.*, "Indoor trajectory reconstruction using mobile devices," in *2018 IEEE 32nd international conference on advanced information networking and applications (AINA)*, IEEE, 2018, pp. 550–555.

[7] Y. Wang, Q. Ye, J. Cheng, *et al.*, "Rssi-based bluetooth indoor localization," in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, IEEE, 2015, pp. 165–171.

[8] F. Bao, S. Mazokha, and J. O. Hallstrom, "Mobintel: Passive outdoor localization via rssi and machine learning," in *2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2021, pp. 247–252. DOI: 10.1109/WiMob52687.2021.9606338.

[9] J.-H. Kim, J.-H. Seong, Y.-S. Ha, *et al.*, "Improved adaptive smoothing filter for indoor localization using rssi," *Journal of the Korean Society of Marine Engineering*, vol. 39, no. 2, pp. 179–186, 2015.

[10] H.-S. Ahn and W. Yu, "Environmental-adaptive rssi-based indoor localization," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 4, pp. 626–633, 2009.

[11] M. V. Barbera, A. Epasto, A. Mei, *et al.*, *CRAWDAD dataset sapienza/probe-requests (v. 2013-09-10)*, Downloaded from https://crawdad.org/sapienza/probe-requests/20130910, Sep. 2013. DOI: 10.15783/C76C7Z.

[12] T. I. Chowdhury, M. M. Rahman, S.-A. Parvez, *et al.*, "A multi-step approach for rssi-based distance estimation using smartphones," in *2015 International Conference on Networking Systems and Security (NSysS)*, 2015, pp. 1–5. DOI: 10.1109/NSysS.2015.7042942.

[13] Y. Zhou, "An efficient least-squares trilateration algorithm for mobile robot localization," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 3474–3479. DOI: 10.1109/IROS.2009.5354370.

[14] K. Calder, "Statistical inference," *New York: Holt*, 1953.

[15] S. Kun, W. Ping, and L. Yingze, "Path loss models for suburban scenario at 2.3ghz, 2.6ghz and 3.5ghz," in *2008 8th International Symposium on Antennas, Propagation and EM Theory*, 2008, pp. 438–441. DOI: 10.1109/ISAPE.2008.4735242.