On the Representativeness of Wi-Fi Data Collection

Giuliano Prestes Fittipaldi⋆, Anne Fladenmuller⋆, Rodrigo de Souza Couto⋄,
Luís Henrique Maciel Kosmalski Costa⋄, and Marcelo Dias de Amorim⋆

⋆Sorbonne Université, CNRS, LIP6
⋄GTA/Poli/COPPE – Universidade Federal do Rio de Janeiro

***Abstract*—Wi-Fi datasets play a crucial role in wireless networking research. They are often the result of extensive, demanding measurement campaigns. Unfortunately, researchers lack a clear assessment of *where* to place probes and *when* they have collected enough data and still obtain a representative view of the environment. Our goal is to make this process more efficient while preserving its rigor. We propose a framework that incorporates a calibration phase to evaluate the representativeness of a dataset. To this end, we use Earth Mover's Distance (EMD) as a similarity metric to quantify data distribution differences and avoid redundant data captures. Through experimental campaigns in three distinct environments, we demonstrate that achieving a significant reduction in data collection effort is possible without compromising measurement reliability.**

***Index Terms*—Wi-Fi, data collection, experiment design.**

I. Introduction

Robust data collection is a cornerstone of scientific research, engineering, and technological development [1]–[3]. The quality and scope of datasets directly influence the accuracy and reliability of experiments. Data collection often constitutes a significant part of the research effort in many fields, including mobile communications, signal processing, and machine learning applications. This paper focuses on link-quality Wi-Fi data collection and seeks to answer the following question: *How to collect Wi-Fi data with reduced effort while keeping the rigor necessary for reliable experiment conclusions?*

Collecting datasets often requires substantial resources, including time to run the experiments, potentially expensive equipment, and personnel effort. Such hurdles often prevent researchers from conducting experiments multiple times or testing various conditions. As a result, many of them unthinkingly restrict data collection to a narrow set of scenarios, which prevents the generalization of the findings. This limitation raises an important question: *Can one collect high-quality datasets in a shorter time frame and/or with fewer collection points?* We explore this question in the context of Wi-Fi packet captures, where the goal is to achieve reliable results while avoiding extensive and time-consuming data collection.

Wi-Fi data collection typically relies on devices that capture Wi-Fi packets, playing the role of probes. Nevertheless, the effectiveness of this probing process depends on various factors, including the chosen distances for placing probing devices and the amount of data captured. Traditionally, Wi-Fi data collection efforts aim to maximize the dataset by collecting data at distances as fine as possible and running experiments for extended durations. Although thorough, this approach may not always be necessary to achieve the desired outcome, particularly when a balance between resource use and data quality is required.

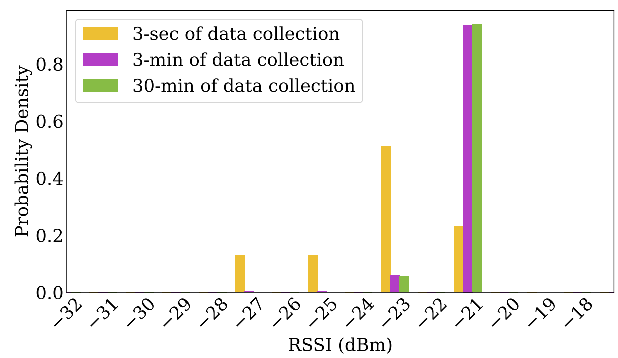


Fig. 1. Distribution of RSSI measurements for 3-second, 3-minute and 30-minute data collection campaigns.

RSSI (Received Signal Strength Indicator) is a good example of Wi-Fi data serving several purposes, such as network performance monitoring, localization refinement, handover assistance, interference detection, and energy optimization at devices. However, RSSI values exhibit significant fluctuations due to the varying channel conditions, making individual measurements unsuitable for characterizing the channel [2]. Instead, a more representative approach involves analyzing the distribution of RSSI values over multiple measurements.

Fig. 1 presents the RSSI distributions from data collection at the same distance X from the source but for three different durations: 3 seconds, 3 minutes, and 30 minutes. With 3 seconds of data collection, the observed RSSI values exhibit noticeable fluctuations. As the measurement duration extends to 30 minutes and more packets are captured, the RSSI distribution converges toward a more stable representation of the underlying signal characteristics. However, while 30 minutes of data collection may seem comprehensive, it can be an *overkill* compared to 3 minutes of data. By analyzing these distributions, we can see that a more targeted approach to data collection – one that carefully balances time and information – can be more efficient while still providing the necessary level of detail for the experiment.

In this paper, we examine the balance between resource utilization and data thoroughness in Wi-Fi data collection and propose a method to assess the feasibility of reducing data collection while preserving the reliability of results. Instead of relying on single RSSI values, we analyze RSSI distri-

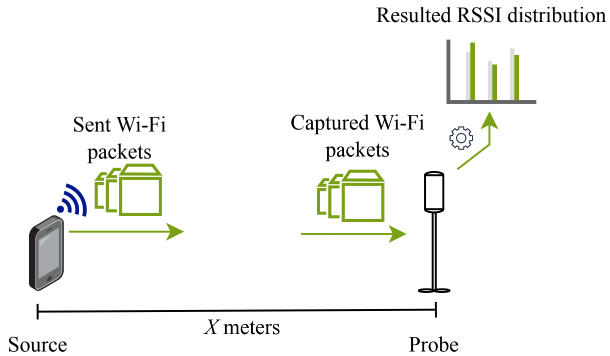


Fig. 2. Typical setup for Wi-Fi data collection.

butions to characterize the wireless channel, using the Earth Mover's Distance (EMD) as a measure of similarity between distributions. Based on EMD values, we apply clustering to identify measurement distances that exhibit similar signal characteristics and determine the amount of data needed at each distance to construct a representative distribution within a given EMD threshold. We evaluate the methodology across three distinct environments – an office, a parking lot, and a campus setting – covering different channel conditions. Our findings demonstrate that comparing RSSI distributions with an appropriate metric enables the grouping of similar distances into clusters, which reduces the number of distinct measurement points needed. We are also able to determine how many Wi-Fi packets we need to capture, in each distance, to build a representative RSSI distribution that meets a specified EMD threshold. Our methodology enables researchers to reduce data collection efforts in an informed manner while maintaining the necessary rigor for experimental reliability. In addition, we offer insights into the refinement of data collection strategies to assist researchers in designing Wi-Fi experiments.

The remainder of the paper is structured as follows. In Section II, we define the problem of Wi-Fi data collection regarding balancing resource constraints while maintaining data quality. In Section III, we discuss the methodology used to evaluate data collection strategies. In Section IV, we instantiate the methodology proposed and analyze the concise data collection over three different scenarios. Section V provides an overview of the related work. Finally, in Section VI, we conclude the article and discuss future research.

II. Problem Statement

Wi-Fi data analyses usually revolve around received signal strength indicator (RSSI) [1], channel state information (CSI) [3], interference patterns [4], and multipath effects [5]. A popular strategy to assess such information is to rely on probe devices (*sniffers*), which passively monitor Wi-Fi traffic. Fig. 2 shows the data collection process, where a probe records packet transmissions from a source X meters apart.

Suppose we want to characterize a wireless channel by measuring the RSSI of Wi-Fi packets. During the data collection campaign, we must decide at which distances from the source we should measure and how many packets each sniffer should capture in each distance to obtain a representative distribution of RSSI values. A common approach is to arbitrarily select probe positions based on intuition and collect data for as long as possible [6]. Nonetheless, this approach may not always be necessary or efficient, as different experiments require different levels of data refinement. The key challenge is collecting data comprehensively while limiting the effort and maintaining experimental rigor.

In real-world scenarios, data collection depends on factors such as the duration of the event of interest, project requirements, and limited access to collection sites. For instance, Barrachina-Muñoz et al. studied Wi-Fi channel bonding inside the Camp Nou stadium during a football match, facing both limited experiment time and restricted access to the environment [7]. The experimental design must thus carefully balance these constraints to ensure data quality and feasibility.

Essential choices in Wi-Fi data collection involve configuring probes and accounting for practical constraints.

Configuration parameters:

- *Sniffer placement*: Determining the positioning of S sniffers to maximize data representativeness.
- *Amount of packets*: Establishing how many packets, N, each sniffer should capture to ensure sufficient information to characterize the environment.

Constraints:

- *Representativeness:* The collected RSSI values should accurately capture the wireless signal patterns, adhering to the level of precision required by the experiment.
- *Data collection effort:*
  - *Sniffer availability:* The available hardware might limit the number of sniffers.
  - *Limited data capture:* Accessibility to the target area, application duration, and sniffers' storage capacity may limit the number of packets a sniffer can capture.

The final goal is to determine the appropriate sniffer placement and number of packets to collect, ensuring that the captured data reflects the precision required for the experiment while considering the data collection effort.

III. Concise Wi-Fi data collection

We propose a methodology that reduces the effort of collecting Wi-Fi data while ensuring the necessary rigor for experimental analysis. The initial phase, conducted only once, considers the finest level of distance resolution and captures the highest volume of packets possible in the experiment to determine the necessary data collection parameters. After this initial step, subsequent experiments – such as monitoring variations across different days – can be conducted more efficiently with fewer probes and a reduced volume of collected data while maintaining reliable results.

A. Metric selection

To characterize the Wi-Fi channel using RSSI distributions, we need a metric that effectively quantifies the differences between these distributions while being appropriate for RSSI measurements. Traditional measures such as Kullback-Leibler (KL) divergence [8] and Jensen-Shannon (JS) divergence [9] are commonly used for comparing probability distributions. Nevertheless, when applied to discrete values, these metrics rely on the degree of overlap between distributions, making them less suitable for Wi-Fi data, where slight RSSI variations may not reflect significant differences in channel conditions.

For example, consider two data collection points, A and B, in which the resulting RSSI distributions correspond to Dirac delta functions $\delta(x - A)$ and $\delta(x - B)$. If $A = -20$ dBm and $B = -21$ dBm, they reflect nearly identical channel conditions, yet KL and JS divergences treat them as entirely distinct due to their lack of overlap. Conversely, if $A = -20$ dBm and $B = -70$ dBm, these metrics still consider them equally dissimilar, failing to reflect the meaningful difference in channel quality.

To address these limitations, we employ the Earth Mover's Distance (EMD) [10], which measures the minimum effort required to transform one distribution into another. Unlike divergence-based metrics, EMD considers both magnitude and distance, making it more suitable for comparing RSSI distributions. Mathematically, EMD is an efficient metric in transportation problems [11], where the goal is to minimize the cost of shifting probability mass from one distribution to another. By using EMD, we ensure a better comparison between measurement positions to reflect real-world signal variations, enabling more informed decisions on data collection.

The EMD is defined as the solution to the transportation problem [11], a bipartite network flow problem, which minimizes the total transportation cost:

$$\text{EMD}(P, Q) = \min \sum_{i,j} c_{ij} f_{ij}, \tag{1}$$

where c_{ij} represents the cost (distance) of moving mass from element i in distribution P to element j in distribution Q, and f_{ij} represents the flow between i and j. The EMD minimizes the total cost of transporting mass between the distributions, and the resulting value is used as the similarity measure.

B. Calibration phase

In this phase, we conduct a single experiment with the maximum data collection effort we intend to make, aiming to obtain the most detailed data possible. The idea is to collect as many packets as possible in the finest feasible granularity regarding probe position. This provides a baseline for comparison with reduced data collection strategies, ensuring we can evaluate the impact of reducing the data collection process for the subsequent experiments.

*1) **Collection distances**:* We use the calibration phase data as a baseline to identify the distances between the probe and the source that are representative of the experimental environment. We use EMD (see Section III-A) as the metric to compare the RSSI distributions across different measurement distances. The goal is to determine whether we can spatially quantize the data collection by clustering distances that produce similar distributions during the calibration phase.

To achieve this, we compute the pairwise EMD between all measurement positions from the calibration phase, forming an EMD matrix where each element $[i, j]$ represents the EMD between the RSSI distributions of distances i and j. While numerous clustering algorithms exist, in this work, we employ a greedy, sequential clustering approach [12]. This method iteratively groups elements, ensuring that each new addition to a cluster satisfies the desired maximum EMD with all previously included elements.

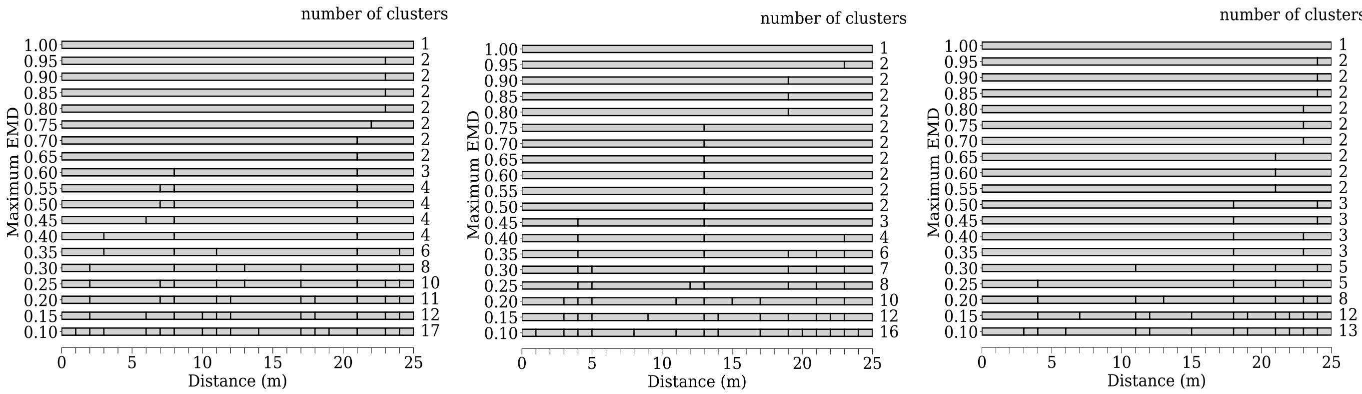
*2) **Dataset volumes**:* Another important consideration is, at each distance, how much data we should capture to reach a certain level of reliability. We analyze the maximum divergence in RSSI distributions for a given packet count using a sliding window approach with EMD. Let D be the dataset from the calibration phase, consisting of the RSSI measurements, where $|D| = S$ denotes the total number of samples.

1) **Window selection**: For each position, we define a window of size W with $W < S$, representing the number of packets we want to evaluate. If this number is sufficient for reliable data, we aim to reduce the data collection in that position to this window size. Each window results in a subset $d_{t,W}$ of the dataset D in the step t:
$$D = \{d_{t,W}, d_{t+1,W}, \ldots, d_{t+W-1,W}\},$$
$$\text{where} \quad 1 \le t \le S - W.$$
2) **Sliding window comparison**: We slide this window across the dataset with a step size of 1, comparing every pair of overlapping windows using EMD. For each pair $(d_{t,W}, d_{t',W})$, where $t < t'$, we compute:
$$\text{EMD}(d_{t,W}, d_{t',W}).$$
3) **Computation of maximum EMD**: After iterating through all possible window pairs, we define the maximum EMD for the given window size W as:
$$\text{EMD}_{\max}(W) = \max_{t,t'} \text{EMD}(d_{t,W}, d_{t',W}).$$
4) **Iteration over different window sizes**: We repeat steps 1, 2, and 3 for different values of W to analyze how the maximum EMD varies with various volumes of data.

This process above measures, in a pairwise fashion, the maximum divergence between distributions when capturing W packets for a certain distance separating the source and the probe. In other words, we can limit the data collection to W packets and still obtain an accurate representation of the link – but with less effort.

IV. Evaluation

Having established the methodology, we now instantiate it in an experimental Wi-Fi data collection scenario. By analyzing different environments, we assess the impact of the distance between the source and the probe, as well as data volume reduction, ensuring that the experiment does not require more effort than necessary.



(a) Office.

(b) Parking.

(c) Campus.

Fig. 3. Clustering of data collection positions according to the maximum EMD.

A. Experimental scenarios

We consider three distinct environments with line-of-sight (LoS) conditions. We follow exactly the setup of Fig. 2. For each experiment, the source transmits 400 packets per measuring point during the calibration phase, with a spatial granularity of one meter. The three scenarios are:

- **Office:** An indoor corridor at Sorbonne Université. Measurements were taken during working hours, with nearby offices occupied, introducing potential interference from human activity.
- **Parking:** A basement parking lot at Sorbonne Université, characterized by an isolated environment with no cellular or Wi-Fi connectivity. This setting represents a remote, interference-free scenario.
- **Campus:** An outdoor location in the courtyard between two towers at Sorbonne Université, offering an open-air propagation environment with potential reflections and diffraction effects from surrounding structures.

B. Representative distances

We begin the methodology instantiation by determining the distance of sniffers in the experimental environments. To this end, we use the EMD metric to cluster measurement distances based on their RSSI distributions, adjusting the maximum EMD threshold to control the allowed intra-cluster divergence. We consider the maximum EMD as a reference value, as the objective is to ensure that each cluster implies a single measurement point in future experiments.

Fig. 3 illustrates the clustering process. The maximum EMD appears on the y-axis and the measurement positions (in meters) on the x-axis. Each horizontal bar represents a maximum EMD level, ensuring all clusters maintain an EMD below or equal to that threshold. The cluster boundaries are marked with lines within each bar, allowing us to identify which sniffers belong to the same cluster. At the end of each bar, we indicate the number of clusters for the corresponding maximum EMD.

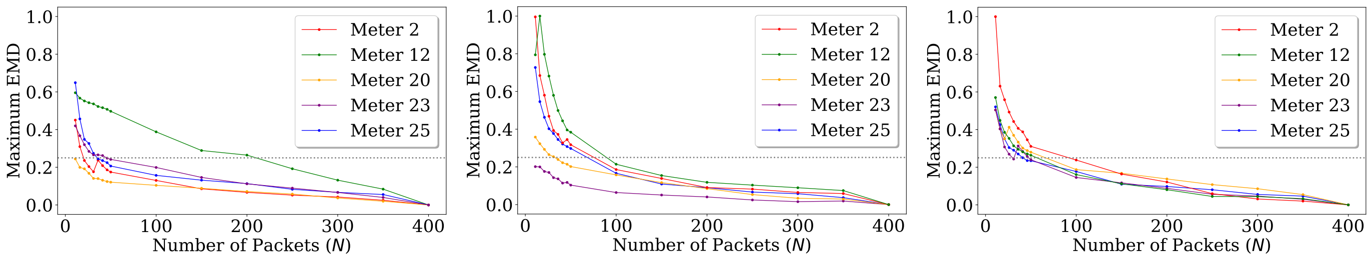
A maximum EMD of one means treating all data collection points as equivalent, leading to a single cluster. If the maximum divergence between data distributions across positions remains within an acceptable error margin, a single measurement may represent the entire environment (a single cluster). Conversely, as we reduce our maximum EMD and require more representative data of each position, the number of clusters increases, as observed in Fig. 3. We do not show the maximum EMD of zero, as it corresponds to the situation where we would have one singleton cluster per measurement position (25 clusters in our case).

We can determine sniffer distance based on experimental constraints. Suppose we have a *hardware* constraint of only five sniffers. In this case, we identify the smallest EMD that results in five clusters (i.e., 0.25) and place one sniffer within each. As an example, for the campus experiment, depicted in Fig. 3(c), this means placing sniffers at positions covering 1–4 meters, 5–18 meters, 19–21 meters, 22-24 meters, and 25 meters. Thus, a representative set of measurement positions would be 2, 12, 20, 23, and 25 meters from the source.

Alternatively, if the constraint is the maximum EMD, and the maximum acceptable EMD is 0.20, we can determine the minimum number of sniffers required to meet this threshold. This would correspond to a minimum of eight sniffers in the campus experiment. It is noteworthy that the data collection requirements vary across different experiments. For the same maximum EMD of 0.20, the office experiment requires 11 measurement points, the parking experiment needs 10, and the campus experiment only requires 8.

C. Data volume to capture

As we continue our methodology, the next step is determining the number of packets to capture at each distance. Fig. 4 illustrates how the maximum EMD (y-axis) varies as a function of the number of captured packets (x-axis) for each evaluated distance. A clear trend is that as the maximum EMD decreases (i.e., the RSSI distribution for a given distance becomes more refined), the number of captured packets increases. This allows us to adjust the data collection process based on the desired level of reliability. For instance, in the campus experiment, based on the sniffer positions chosen in the previous step (at



(a) Office. (b) Parking. (c) Campus.

Fig. 4. Maximum EMD vs. Number of packets used in the experiment.

2, 12, 20, 23, and 25 meters), by analyzing the intersection of the gray dotted line with the curves, as illustrated in Fig. 4(c) for our example, we can determine the minimum number of packets necessary to capture in each position to achieve the same reliability in the measurements with a maximum EMD of 0.25. In our example, the number of packets to collect in the campus experiment is 50 for 2m, 35 for 12m, 40 for 20m, and 20 for both 23m and 25m. Following our approach, for the chosen maximum EMD threshold, the total number of collected packets across all distances amounts to 165.

This methodology not only assists our decision-making regarding probe distance and the volume of data to capture but also provides insights into the data capture process. By analyzing Fig. 4, we notice that the parking experiment has a more consistent behavior throughout different data volumes. This observation might come from the fact that the parking area is on the -5th floor of the campus, where interfering signals are almost absent. Conversely, in the office experiment, a distance of 12 m presents itself as an outlier, requiring 6 to 40 times more packets than other distances to achieve the same data reliability of 0.25 EMD. This phenomenon could indicate an artifact at that position that results in more inconsistent measurements or even propagation effects related to that distance.

D. Capture reduction and representativeness

Table I shows a more detailed analysis of the packet volume reduction. For each selected maximum EMD value, the table shows the average number of packets required (in the "average packets" column) and the corresponding ratio Δ compared to the baseline capture. For instance, in the Office scenario, with a maximum EMD of 0.2, only 24.75% of the initially planned packets need to be captured, reducing the data collection effort by a factor of four. This implies that the time allocated for a single experiment in the office could be reallocated to conduct four separate experiments. It also raises an important issue regarding what representativeness is in experimental work.

Representative data collection involves a trade-off between data precision and experimental scope. With limited efforts, researchers must choose between collecting accurate measurements in a few selected scenarios or gathering broader data that captures the full range of real-world conditions.

For example, consider deploying Wi-Fi in a concert arena with limited access. Measuring wireless characteristics only

TABLE I
AVERAGE VOLUME OF PACKETS CAPTURED AND VOLUME REDUCTION, PER MAX EMD IN THE THREE SCENARIOS.

	Office		**Parking**		**Campus**	
Max EMD	**Average Packets**	**Δ Volume**	**Average Packets**	**Δ Volume**	**Average Packets**	**Δ Volume**
1.0	10	2.50%	11	2.75%	10	2.50%
0.9	10	2.50%	11	2.75%	11	2.75%
0.8	10	2.50%	11	2.75%	11	2.75%
0.7	10	2.50%	15	3.75%	11	2.75%
0.6	11	2.75%	17	4.25%	12	3.00%
0.5	11	2.75%	20	5.00%	17	4.25%
0.4	32	8.00%	25	6.25%	24	6.00%
0.3	47	11.75%	56	14.00%	47	11.75%
0.2	99	24.75%	93	23.25%	110	27.50%
0.1	230	57.50%	200	50.00%	230	57.50%

at the stage or a few specific rows might yield high-precision measurements but miss critical spatial variations. In contrast, sampling data from all seating areas – even if individual measurements exhibit slightly higher variability – offers a more comprehensive picture of network behavior by accounting for obstacles, interference, and other dynamic environmental factors.

Moreover, the ideal balance between precision and scope depends on the nature of the system under study. A broader scope in data collection provides more robust generalization in highly dynamic environments where conditions fluctuate over time or regarding space. On the other hand, increased measurement precision may be more valuable in static or highly controlled settings. By shifting the focus from absolute precision to the representativeness of collected data, researchers can ensure that experimental results are reliable and applicable to real-world scenarios.

V. Related work

The use and analysis of Wi-Fi data covers a significant slice of wireless communications studies. Many of the works aim to provide a basis for application and thus rely on experimental analysis and data collection.

As an example of experimental wireless communication work, Bertier et al. present an empirical characterization of high-speed device-to-device (D2D) communication technologies in Android, specifically Wi-Fi P2P and Nearby Connections [13]. The data collection was performed with high precision, measuring the goodput at intervals of 1 meter up to 10 meters, then every 5 meters up to 100 meters, and finally every 10 meters up to 300 meters. While the data collection campaign was thorough, it might have been possible to reduce the effort, allowing more time to explore different scenarios, or adjust the collection parameters to better suit the requirements of the different collection distances.

Some localization studies also rely on Wi-Fi data collection, where dataset development depends on decisions regarding sensor placement and the amount of data gathered at each point. Ma et al. investigate a Wi-Fi indoor positioning system (IPS) based on the IEEE 802.11mc fine-timing measurement (FTM) scheme, also known as Wi-Fi RTT [14]. Using a commercial smartphone and Wi-Fi access points, they conduct experiments in real-world indoor environments to assess range performance and enhance positioning accuracy. Their dataset consists of measurements from a 12-point grid, with data collected at 0.5-meter intervals up to 20 meters under different channel bandwidths (20, 40, and 80 MHz). At each position, measurements were recorded for five minutes at the highest available scanning frequency.

The idea of reducing data collection efforts while preserving data quality has been explored in other experimental fields but remains largely unaddressed in Wi-Fi data collection. An Approximate Data Collection (ADC) approach for sensor networks was proposed by Wang et al. which reduces communication costs while maintaining predefined error bounds [15]. Their method clusters sensor nodes, exploits local data correlations, and performs global data approximation at the sink node. However, their approach still requires deploying all sensors at maximum granularity, as data from all nodes must be collected before applying their optimization, meaning the overall deployment effort remains unchanged.

In contrast, our approach focuses on reducing the deployment effort itself by incorporating a calibration phase. This aims to minimize unnecessary time consumption, equipment costs, and personnel mobilization for Wi-Fi data collection experiments.

VI. Takeaways and Future Work

While our methodology provides guidance for efficient Wi-Fi data collection, we do not suggest avoiding rigorous, exhaustive measurements when necessary. The goal is not to reduce data collection indiscriminately but to refine it based on experimental constraints and environmental factors. We do not advocate universally fixed thresholds either, as each experimental scenario requires a careful calibration phase, as proposed, based on its specific constraints and objectives.

Primarily, our methodology demonstrates that Wi-Fi data collection does not have to rely on arbitrary decisions. By incorporating a calibration phase and a well-chosen metric, researchers can apply their data collection efforts more efficiently, enabling the exploration of additional scenarios within the same time and equipment constraints. Since Wi-Fi signal propagation varies across different LoS conditions – as expected – our methodology allows researchers to tailor data collection efforts, ensuring that all collected positions adhere to the same maximum EMD and maintain consistent reliability.

Future work could extend this approach by evaluating the impact of reducing data collection efforts within the same scenario and across scenarios that exhibit similar characteristics (e.g., different corridors within the same building). Additionally, we intend to explore alternative clustering algorithms for sniffer placement and investigate clustering behavior at longer distances.

References

[1] A. Natarajan, V. Krishnasamy, and M. Singh, "A machine learning approach to passive human motion detection using wifi measurements from commodity iot devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023.

[2] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, Dec. 2013. [Online]. Available: https://doi.org/10.1145/2543581.2543592

[3] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, Jun. 2019. [Online]. Available: https://doi.org/10.1145/3310194

[4] T. Pulkkinen, J. K. Nurminen, and P. Nurmi, "Understanding wifi cross-technology interference detection in the real world," in *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, 2020, pp. 954–964.

[5] X. Wang, A. Yu, K. Niu, W. Shi, J. Wang, Z. Yao, R. C. Shah, H. Lu, and D. Zhang, "Understanding the diffraction model in static multipath-rich environments for wifi sensing system design," *IEEE Transactions on Mobile Computing*, vol. 23, no. 11, pp. 10 393–10 410, 2024.

[6] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, p. 269–282, Aug. 2015. [Online]. Available: https://doi.org/10.1145/2829988.2787487

[7] S. Barrachina-Muñoz, B. Bellalta, and E. W. Knightly, "Wi-fi channel bonding: An all-channel system and experimental study from urban hotspots to a sold-out stadium," *IEEE/ACM Transactions on Networking*, vol. 29, no. 5, pp. 2101–2114, 2021.

[8] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951. [Online]. Available: https://api.semanticscholar.org/CorpusID:120349231

[9] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[10] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 59–66.

[11] G. B. Dantzig, *Application of the Simplex Method to a Transportation Problem*. John Wiley and Sons, 1951, pp. 359–373.

[12] J. A. Hartigan, *Clustering Algorithms*. John Wiley & Sons, 1975.

[13] C. Bertier, M. Dias de Amorim, F. Benbadis, and V. Conan, "Modeling realistic bit rates of d2d communications between android devices," in *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 315–322. [Online]. Available: https://doi.org/10.1145/3345768.3355918

[14] C. Ma, B. Wu, S. Poslad, and D. R. Selviah, "Wi-fi rtt ranging performance characterization and positioning system design," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 740–756, 2020.

[15] C. Wang, H. Ma, Y. He, and S. Xiong, "Adaptive approximate data collection for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 6, pp. 1004–1016, 2011.