

Efficient Spatial-Temporal Angle-Delay Analysis Scheme for Massive MIMO Indoor Tracking

Van-Linh Nguyen^{*}, Harry Wong Hung-Jun^{*}, Yu-Chia Lin^{*}, Ren-Hung Hwang[†]

^{*}Dept. of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan

[†]College of Artificial Intelligence, National Yang Ming Chiao Tung University, Tainan, Taiwan
{nvlinh, hwhj, aaaee9104}@cs.ccu.edu.tw, rhhwang@nycu.edu.tw

Abstract—Radio positioning is critical for many indoor applications, such as behavioral monitoring and autonomous robots. Mobile users, however, can also be exposed to surveillance risks due to this capability. This work presents a Spatial-Temporal Angle-Delay Analysis Scheme (STADAS) for massive MIMO wireless networks that can help the attacker to track a user without the need to enter buildings. First, we transform the channel state information (e.g., angle of arrival, time of arrival) from massive MIMO transmission gained over time into living Angle-Delay profiles (ADPs) with fixed objects (building walls, furniture) and a moving object (the mobile user). Second, a generative adversarial network learning model is used to remove distorted data points from Angle-Delay video frames. The processed ADPs are trained with a Deep Convolutional Neural Network (DCNN)-based model on estimating the user’s location. Evaluations on an empirical dataset indicate that radio positioning capabilities in emerging wireless communication technologies such as mmWave MIMO can pose severe privacy and surveillance threats.

Index Terms—Wireless security, User tracking, Radio-based Localization.

I. INTRODUCTION

Massive MIMO and directional transmission techniques are expected to become the dominant antenna design and data transmission technologies for video and high-bandwidth applications in 5G wireless networks. By training on beam signal space and exploiting spatial-temporal signal processing, massive MIMO enables high-accuracy localization [1], [2]. For example, an accurate location is used in indoor activity trackers, geofencing for people with dementia, and augmented reality [2]. The radio positioning feature will be critical in indoor environments where the Global Positioning System (GPS) often performs poorly.

However, radio-based positioning technology can be potentially abused for user surveillance. In this work, we develop an efficient scheme that can track a mobile user accurately, even without going into the user’s building. The tracking information can reveal the user’s movement trajectory in a real-time manner. This raises new concerns about user privacy risks in one of the most promising antenna technologies. As illustrated in Figure 1, an attacker may deploy a tracking device (e.g., fake base station) to infer cellular/WiFi signals and track the target user in a restricted access building. Another typical example of harmful tracking is when the adversary tries to follow the VIP targets (while they are using their phones).

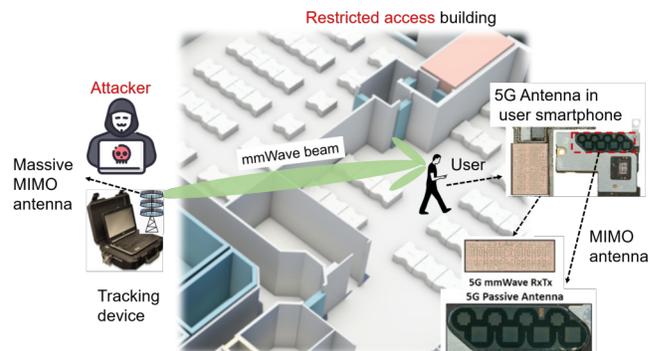


Fig. 1. The illustration of exploiting signal spatial-temporal signal processing of directional massive MIMO-based technologies to track a user in a restricted access building without physically going into.

A. State-of-the-art studies

Indoor tracking has been an important research topic for years. Currently, the research community is making a lot of efforts to realize a long-awaited goal of centimeter tracking accuracy in various environments, e.g., 5G mmWave [2]–[5]. There are five main approaches of indoor tracking algorithms. The first approach is a set of triangulation and trilateration-based methods, where uplink and downlink Angle of Arrival (AoA) or Time Difference of Arrival (TDOA) measurements are used to estimate the user’s position [1]. However, the weakness of this technique is at least three base stations must be available to resolve the user’s relative location. Meanwhile, its accuracy error can be up to several meters. The second approach highlights cooperative localization [6] where the tracking accuracy can be enhanced by exploiting peer-to-peer feedback among the nearby users for estimating a specific user’s location [7]. However, it is ideal to assume that the users are always in communication. Moreover, in dynamic networks, the topology is constantly changing. Due to this, it is difficult to establish a stable network state for launching localization.

In the third technique, machine learning is the most common model to be used. The machine learning-based methods often exploit the advantage of large-scale training on channel state information (CSI) fingerprint (e.g., amplitude, angle, Radio Signal Strength Indicator (RSSI)) datasets to identify the user location from the ground truth patterns [5], [8]. However, although this technique can enable high accuracy, the require-

ments of large-scale fingerprinting datasets can be a challenge in unknown structures. Further, there is insufficient data to collect in many cases (e.g., sparse user density). The fourth type of tracking is multipath-based [1], where AoA/TDoA values are extracted from multiple path components. However, this technique’s accuracy significantly degrades when multiple path sources are detected in close proximity. Finally, many studies have relied on the hybrid tracking approach, i.g., using the combination of multiple positioning methods and data fusion, e.g., Bayesian [7]. In order for a hybrid approach to achieve high accuracy, a variety of data sources must be used.

B. Contributions

This work presents a novel approach, namely Spatial-Temporal Angle-Delay Analysis Scheme (STADAS), to demonstrate high-accuracy indoor tracking. In the first step, a deep learning model is proposed to run on the Angle-Delay profiles (ADPs), which are extracted from propagation models and path loss statistics of well-known material penetration, to estimate the relative location of a user. The estimations are then used for trajectory generation with the help of state-of-the-art video prediction technology. The results demonstrate that our scheme can yield high-accuracy indoor surveillance capability, even without physical intrusion into the target environment.

II. SYSTEM MODEL AND PROBLEM STATEMENT

This work aims to track a specific mobile user in a department building. We assume that adversaries can deploy an anchor (truck base station) in known positions to locate a single user. Suppose that the user’s location is unknown. The user uses cellular mmWave networks or WiFi mmWave networks. The antennas of the anchor and the user are Uniform Linear Array (ULA) type. The anchor and the user have N_R and N_T antennas, respectively. The mmWave system uses orthogonal frequency division multiple access (OFDM) with K subcarriers.

A. Tracking attack model

In this work, we assume that the attacker can passively collect the user’s signals by deploying eavesdropping devices around the building. In wireless networks, this eavesdropping attack is common and used, especially against directional wireless communications such as mmWave [9]. The other way is that the attacker can launch jamming attacks against genuine base stations. The user is forced to broadcast signals in order to find the nearby base station (to re-establish the connection). In this case, the attacker’s tracking device attracts the user connection and acts as a fake base station. Since radio-based tracking takes place at the physical layer, tracking is possible once the transmission and signals exist. Due to this, the attacker can track the user even if there are errors during data transmission, such as authentication errors.

B. Signal and communication model

Similar to [8], we adopt a wideband geometric mmWave MIMO channel with C clusters. Each cluster can constitute up to L rays/paths between the anchor (receiver) and the user (transmitter). Then the channel impulse response (CIR) matrix for each sub-carrier k ($k = 1, 2, \dots, K$) at the time t is given by

$$h[k](t) = \sqrt{\frac{N_R}{\rho_{pl}}} \sum_{c=1}^C \sum_{l=1}^L \alpha_{c,l} a_{c,l}^R(\phi) a_{c,l}^T(\theta) e^{-j2\pi \frac{\tau_{c,l}}{kT_s}}, \quad (1)$$

where ρ_{pl} is the path loss. T_s and $\tau_{c,l}$ denote the sampling period and the delay belonging to the l th path of the c th cluster, $\alpha_{c,l}$ is the complex channel gain [10]. ϕ and θ denote the physical angle of arrival (AoA) and angle of departure (AoD) of each cluster ($\phi \in [\frac{\pi}{2}, \frac{3\pi}{2}]$ - right unit circle, $\theta \in [\frac{3\pi}{2}, \frac{\pi}{2}]$ - left unit circle). $a_{c,l}^R(\phi)$ denotes the array response vector at the anchor from the l th path and the c th cluster. $a_{c,l}^R(\phi)$ is given by:

$$a_{c,l}^R(\phi) = \frac{1}{\sqrt{N_R}} [1, e^{j2\pi d \sin(\phi)}, \dots, e^{j2\pi(N_R-1)d \sin(\phi)}]^T, \quad (2)$$

where d is the distance between two elements of the anchor antenna array (as illustrated in Figure 1) in wavelength.

Suppose that the signals penetrate via multiple walls. In this work, we use the Wall Attenuation Factor model as described in [11]. Accordingly, the signal path loss ρ_{pl} in decibels (dBm) caused by obstacle attenuation between the transmitter and the receiver is modeled by

$$\rho_{pl} = \begin{cases} \rho(d_0 f_c) - 10n \log(\frac{d_{au}}{d_0}) - W * F & \text{if } W < M \\ \rho(d_0 f_c) - 10n \log(\frac{d_{au}}{d_0}) - M * F & \text{otherwise,} \end{cases} \quad (3)$$

where $\rho(d_0 f_c)$ is the signal power at some reference distance d_0 of the carrier frequency f_c , e.g., $d_0 = 1m$. n denotes the path loss exponent that increases with distance. d_{au} is the Euclidean distance between the anchor and the user. M is the maximum number of obstructions (walls), a constant. W is the estimated number of obstructions between the anchor and the user. F is the wall attenuation factor. Both n and F depend on the building layout and construction material. The values of signal loss in several wall materials can be found in [12], [13]. Finally, the general channel impulse response matrix between the anchor and the user at the time t can be written as follows:

$$H(t) = [h[1](t), \dots, h[k](t), \dots, h[K](t)]^T, \quad (4)$$

C. Problem statement

To take advantage of the rich deep learning techniques, we need to transfer the CIR matrix H into a linear form, the so-called Angle-Delay profile G . G is estimated by mapping the space-frequency domain to the angle domain and the delay domain at the time t as follows:

$$G(t) = V^H H(t) F = \frac{1}{\sqrt{N_R}} e^{-j2\pi \frac{z(q - \frac{N_R}{2})}{N_R}} H(t) \frac{1}{\sqrt{K}} e^{-j2\pi \frac{zq}{N_R}}, \quad (5)$$

where $V^H (\in C^{N_R \times N_R})$ and $F (\in C^{K \times K})$ are two discrete Fourier transform matrices, (z, q) element represents the complex gain associated with the z th AoA and the q th ToA [14]. Now the problem of locating the user from the channel impulse response matrix H is equal to finding the user position by running deep learning algorithms on Angle-Delay channel profiles G . In the following section, we present a novel scheme with two deep-learning models and an averaging function to enhance user tracking accuracy.

III. PROPOSED SPATIAL-TEMPORAL ANGLE-DELAY ANALYSIS SCHEME FOR INDOOR TRACKING

This section introduces a novel scheme to enhance angle-delay-based tracking accuracy by correlating potentially inconsistent data points from Fourier transformation and averaging location estimation results after multiple trials. The proposed data point correlation and location estimation calibration mechanism are efficient for enhancing tracking performance, particularly when the user repeatedly moves on the same route.

A. Building Angle-Delay Data Points profile

The channel state information must be transformed into a multimedia representation to exploit video prediction techniques. In this work, we use an Angle-Delay profile (ADP). According to [14], ADP represents the angle of arrivals of the received signals in terms of the delay with respect to the arrival paths in multi-path transmission. Also, Angle-Delay profiles preserve the user's motion when stacked temporally. The channel state information is first collected during the communication period as illustrated in Figure 2a). The information is then transformed into corresponding Angle-Delay profiles through linear Fourier transformation as illustrated in Figure 2b). Figure 2c) shows the data points of (angle, delay) information from Angle-Delay profiles over time of each user (if there are multiple users on the signal map collection). The time-based data points are combined to build a video where static objects like building walls are fixed lines and dynamic objects representing user movements are dot images as illustrated in Figure 2d). However, the video's data points can be distorted due to noise and signal inference. The following subsection details how we use a state-of-the-art video frame prediction technique to correlate distorted Angle-Delay data points.

B. Frame Prediction for Correlating Angle-Delay Data Points

In an environment with many obstacles, e.g., a hotel, signal-based localization accuracy is negatively impacted by signal scattering or multi-source overlap. As a result, transformed angle-delay data points G can be distorted [8]. Several state-of-the-art deep learning techniques [15] have exploited frame prediction to correlate "distorted" data points. Assume that we have a sequence of accurate ADP records collected at the time t (e.g., after T_s seconds of tracking the user on the building hall), $G(t) = \{G(t_1), G(t_2), \dots, G(t_{T_s})\}$. In this work, we use a Spatial-Temporal Multi-Frequency Analysis Network (STMFANet) [16] to predict the next frame of the sequence.

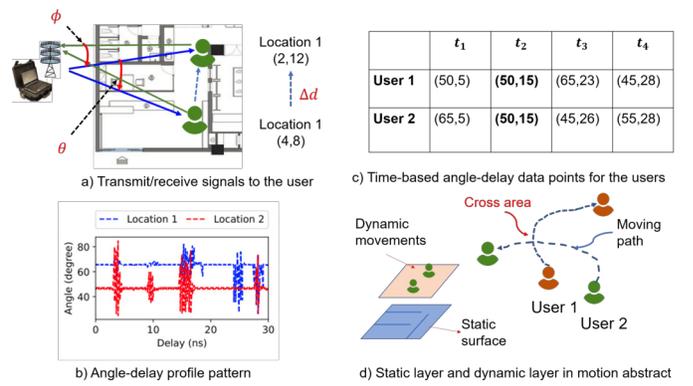


Fig. 2. The illustration of cross areas of the moving paths of two users. The signals of a user can be wrongly labeled to the others at the cross area.

STMFANet is a generative adversarial network where the Generator and the Discriminator are trained to predict the next video frames. The loss \mathcal{L}_{pre} of the prediction network consists of two losses (image domain loss \mathcal{L}_{IMG} and adversarial loss \mathcal{L}_{ADV}) as follows:

$$\begin{aligned} \mathcal{L}_{pre} &= \lambda_1 \mathcal{L}_{IMG} + \lambda_2 \mathcal{L}_{ADV} \\ &= \lambda_1 \left(\sum_{i=1}^{T_s} \|G(t_i) - \hat{G}(t_i)\|_2^2 + \Psi \right) - \lambda_2 \log D([G(t), \hat{G}(t)]), \end{aligned} \quad (6)$$

where λ_1 and λ_2 are hyper-parameters to trade-off between two distinct losses, Ψ is the Gradient Difference Loss [16].

C. Deep CNN-based User Location Estimation on Angle-Delay Data Points

The final task is to locate the user. We run a Deep CNN-based User Location Estimation (DCULE) on the "correlated" data points. DCULE consists of m CALP modules and softmax function. m is a constant, e.g., $m = 2$. Each CALP module includes four layers [14]: (1) a convolution operation layer with $N_R \times K$ convolutional Kernels; (2) an activation function layer (e.g., RELU); (3) a local response normalization layer; (4) a pooling layer. The input for the first layer of CALP includes correlated data points at the time t . The output after softmax is the estimated locations of the user at the time t is $x(t) = \{x(t_1), x(t_2), \dots, x(t_{T_s})\}$. In other words, $x(t)$ are the results of running DCULE on $\hat{G}(t)$. The final loss of running DCULE module is bounded by the cross-entropy loss function as defined as follows:

$$\mathcal{L}_{est} = \frac{1}{D} \sum_{i=1}^D x'(t) \log(x(t)), \quad (7)$$

where $x'(t)$ and $x(t)$ denote the probability distribution of labeled training data (past locations) and the DCNN training's output. D is the number of training episodes.

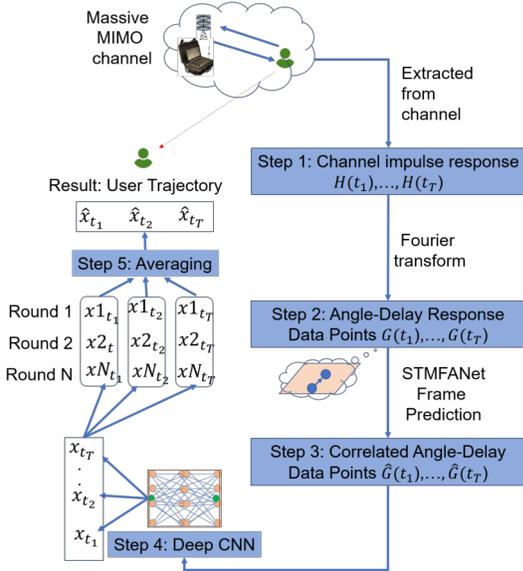


Fig. 3. The illustration of the proposed indoor positioning workflow.

D. Multiple-round-based User Monitoring and Averaging Location Estimation

One of the typical surveillance cases is that the tracker does not have the drawing of the building or the inside structure. While revealing the in/out routes and a relative location of the target objects is still possible, knowing the exact room/place is complicated. To enhance the results for this blind tracking, the building architecture for ray tracing can be roughly estimated by the standard sizes of several structures. However, the furniture’s effect on signal loss can proportionally impact the tracking performance. To gain the best results, we can run N rounds of DCULE-based user tracking with different locations of the anchor (e.g., the attacker moves around the building as illustrated in Figure 1) or the cooperation of several agents. And the target user location $\hat{x}(t)$ is calculated by averaging the estimated locations of N round trials, i.e., $\hat{x}(t) = \frac{1}{N} \sum_{j=1}^N x_j(t)$. Figure 3 summarize the workflow of processing steps in the proposed system.

A building drawing, if any, can significantly help refine the tracking results. Specifically, if there are outlier data points from the estimated locations (after step 4 in Figure 3), they can be easily detected by referring to a grid of data points from the walls and rooms’ coordinates. Without outlier data points, the estimated locations after averaging (after step 5 in Figure 3) with knowing the building structure can be more accurate than those without knowing. Further, the data points can be checked via ray tracing techniques [2], given the known geometric structure of the building.

IV. EVALUATION RESULTS AND ANALYSIS

We evaluate the performance of our tracking mechanism via simulation of indoor “I3” scenario in DeepMIMO dataset [17]. The dataset is scaled for a $100m \times 110m \times 3m$ building. Also, as for the propagation model, each channel path can

TABLE I
THE TRAINING HYPERPARAMETERS AND NETWORK CONFIGURATION

Parameter	Value	Parameter	Value
Antenna	ULA	Carrier frequency	2.4/60GHz
No of episodes D	200	Bandwidth	20MHz
N_R/N_T	32-128/16-128	No of subcarriers K	32
Learning rate	1-e4	Sampling period	2ns
Activation function	ReLU	Optimizer	Adam

undergo a maximum of 2 reflections and 1 transmission before reaching the receiver. We also use the path loss and phase change for RF propagation ray in MATLAB to trace the signals when the anchor moves around the same building. The user randomly walks with a maximum speed of $1.3m/s$ ($5km/h$). The sampling period T_s is at $2ns$. The operating frequencies are 2.4 GHz and 60 GHz, as set in the dataset. The anchor (spy agent) is outside the building to avoid the target’s suspicion by setting a negative location value. The tracker can move around the left and the bottom side of the building at a distance of $-20m$ to $-30m$. The other parameters for training hyperparameters and network configuration are summarized in Table I. We compare the accuracy performance of the proposed method with three state-of-the-art tracking models (whose codes are available): (1) DyLoc [8]; (2) LEAP [3]; (3) DCS-RADAR [5], [18].

At the default configuration (as shown in Table I, $N_R = 32$), Figure 4(a) shows the cumulative distribution of the estimation errors for four methods. Accordingly, the proposed method achieves the best performance with 95.6% reliability in tracking accuracy below one meter. By exploiting the power of deep recurrent neural networks (e.g., PredRNN), DyLOC ranks second in tracking performance with 91.3% reliability for lower-1-meter accuracy. The positive result of the proposed approach over DyLOC comes from the advantage of the STMFANet mechanism in overcoming distorted frames. STMFANet encourages two competitors in learning (Generator and Discriminator) that can correlate the distorted data points much faster than the long-term and hierarchical-memory learning model did (i.e., PredRNN in DyLOC). Further, in our tracking scheme, the outlier data points can be mitigated through the built-in multiple-round-based monitoring and position averaging mechanism, which DyLOC did not support. Lacking calibration mechanisms for outlier data points also makes the two remaining methods’ performance ranks last. Besides, we found that trajectory smoothing via the Kalman filters in LEAP or Particle Filter in DCS-RADAR performs poorly if the target user randomly walks or the signals are heavily distorted (that causes many outlier data points). Finally, due to exploiting spatial-temporal features, memory-based learning models like ours and DyLOC can give better overall accuracy when training on time-series data points than the spatial-based learning models (LEAP, DCS-RADAR) do.

Our multiple-round-based position monitoring and averaging module are particularly useful when the building structure is complicated, or the environment is noisy (many thick walls or furniture exist). For example, as shown in Figure 4(b), by

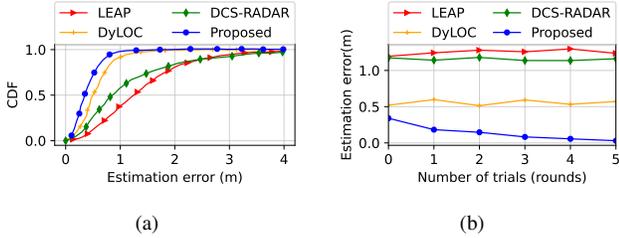


Fig. 4. The localization performance comparison of the proposed method and three state-of-the-art tracking models: a) CDF of the estimation errors; b) Estimation error over the number of trials.

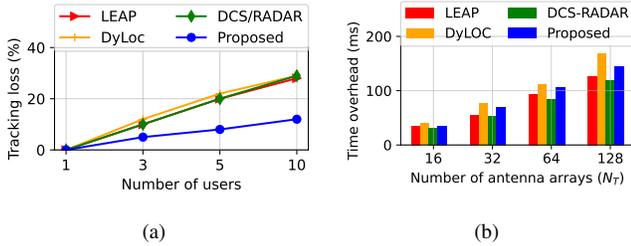


Fig. 5. The performance comparison of the proposed method and three state-of-the-art tracking models in a) Tracking loss over the number of trials; b) Time overhead with the corresponding number of antenna arrays configuration;

running multiple trials (three rounds with different locations of the anchor) and averaging the estimated locations, the proposed method can significantly reduce the location estimation error to several centimeters. By contrast, without a strong spatial-temporal outlier data point correlation and multiple-trial estimation mechanism, the tracking performance of all three methods (DyLOC, LEAP, DCS-RADAR) remains no significant enhancement, even after many trials.

In the worst case, there are multiple mobile users in the building, and the users may cross-walk each other. As shown in Figure 5(a), the ratio of tracking loss on a specific user increases for all methods due to the challenge of distinguishing two too-close signal sources or signal angle overlaps. In this case, the proposed method shows a slight improvement with 10% less loss than the other methods. The positive enhancement comes from the well-correlated data points after averaging user locations from running 3-rounds of our algorithm at the different locations of the anchor. However, due to training on large spatial-temporal data point spaces, as shown in Figure 5(b), the proposed tracking approach suffers a little bit more time overhead, i.e., delay in responding to the localization requests, compared with LEAP and DCS-RADAR but lower than DyLOC does. The longest time overhead is when the transmitter (user) uses a device with large antennas that accidentally creates many ADP inputs for STMFANet and DCULE training. DyLOC suffers the longest time overhead due to the higher computation expenditure of the PredRNN-based learning model in DyLOC over the STMFANet-based

TABLE II
AVERAGE RMSE COMPARISON OF THE TECHNIQUES WITH A SENSITIVITY ANALYSIS

f_c	ρ	M	W	P	Material	LOS	RMSE (m)			
							Leap	Dyloc	DCS-RADAR	Proposed
2.4GHz	0.10	3	2	23	Brick	Yes	1.25	0.59	1.11	0.48
	0.45	3	2	23	Concrete (203mm)	No	1.79	0.82	1.56	0.65
	0.10	3	2	34	Brick	Yes	1.18	0.37	1.05	0.33
	0.45	3	2	34	Concrete (203mm)	No	1.61	0.76	1.49	0.57
	0.10	4	3	23	Brick	Yes	1.48	1.04	1.33	0.51
	0.45	4	3	23	Concrete (203mm)	No	1.95	1.17	1.64	0.94
Avg.							1.54	0.79	1.36	0.58
60GHz	0.10	3	2	23	Brick	Yes	0.72	0.44	0.59	0.15
	0.45	3	2	23	Concrete (203mm)	No	1.85	0.97	1.71	0.69
	0.10	3	2	34	Brick	Yes	0.35	0.13	0.26	0.08
	0.45	3	2	34	Concrete (203mm)	No	1.29	0.63	1.54	0.47
	0.10	4	3	23	Brick	Yes	0.89	0.55	0.73	0.17
	0.45	4	3	23	Concrete (203mm)	No	2.01	1.02	1.98	0.76
Avg.							1.18	0.62	1.13	0.38

model in the proposed method in frame prediction.

For sensitivity evaluation, we found that the proposed tracking method can significantly enhance accuracy in several cases: (1) the high frequency is adopted in wireless communications; (2) Line-of-Sight (LOS) paths exist between the anchor and the user in one of the trials; (3) the target user's transmission power is strong enough. The performance comparison of four techniques with sensitivity analysis is detailed in Table II. Tracking on the 60GHz-based communications and LOS path existence can yield an average of Root Mean Square Error (RMSE) around 0.4m, even with less cooperation of the anchors. In this case, the high directional signals contribute to the high accuracy Angle-Delay profile resolution, particularly the arrival of angle. For the other frequencies (e.g., 2.4GHz), the average estimation error is around 1m. Note that the accuracy of the localization for indoor tracking in this work is significantly lower than the empirical methods or the studies as reported in [1], [2], [11]. This is because the proposed method has no help from the ground truth comparison (i.e., the anchor cannot get in the building without a suspect of the target) to refine the results in the tracking step. The refinement is primarily based on multiple-round-based observation, which neither of DyLOC/LEAP/DCS-RADAR supports.

The multiple-trial observation can also be applied in specific tracking missions, e.g., where the targets are supposed to go in/out of the building several times. Without this mechanism, compromised access points or building cameras are required to refine the tracking state. However, we found that averaging the estimation results from many anchors at different locations can yield unexpected results. The negative influence likely worsens if the anchors' views on the target's location are so different, e.g., due to the influence of overlapped arrival of scattering signals from different orientations. These phenomena often appear in the testing with thick concrete walls or no LOS path (Table II).

When the building structure is thick with many concrete walls (e.g., $W = 3$), we also found that the RMSE per-

formance of channel-based approaches such as LEAP and DCS-RADAR increases rapidly due to noise, particularly for high carrier frequencies (e.g., $f_c = 60GHz$). By contrast, the techniques based on the time difference, such as DyLoc and the proposed method, are less vulnerable to noise ($\rho = 0.45$) but highly vulnerable to multipath propagation (affects the time of phase of arrival). The proposed system can overcome the multipath propagation by exploiting the averaging from multiple anchors and thus suffers less RMSE than the other methods. As shown in Table II, the average RMSE is at $0.38m$ compared to that of $1.18m$, $0.62m$, and $1.13m$ from LEAP, DyLoc, and DCS-RADAR, respectively. Besides, the power transmission of the user (e.g., $34dBm$) in mmWave communications also contributes to reducing RMSE in four models, particularly the proposed technique, given the ease of determining the angle of arrival if the primary beam can be detected. Finally, Table II also shows that the user's estimated location accuracy can be negatively impacted if there is no LOS path between the target user and the anchor. For example, the proposed method suffers an average of $0.15m$ in RMSE if a dominant LOS path exists (the first row of the testing with $f_c = 60GHz$). By contrast, the RMSE soars up to $0.69m$ if there is no such LOS path.

Defending against signal-based tracking methods has been a challenge. Turning off the smartphone and creating well-designed obstacles in the building can help mitigate signal-based tracking threats since AoA/TDoA estimation or power transmission will be weakened. However, the turn-off method seems impractical since it can interrupt the user's connectivity. Another way is to steer main lobe signals into specific locations of authorized user groups. Still, that signal steering approach is vulnerable to eavesdropping attacks (e.g., by using reflectors) if the base station is out of the building (a common case). Besides, given no general structure for the buildings and the challenges of obtaining the ground truth data, a possible extension for this work is to build a calculation model for finding the best locations to place the anchors where the tracking can perform best without multiple trials.

V. CONCLUSION

This work presents an efficient signal-based tracking scheme to follow a specific user in indoor buildings, namely STADAS. The system can reveal the user trajectory through two deep-learning models on the channel state information (e.g., AoA, TDoA) from massive MIMO transmission gained over time. The simulation results show that the proposed method can maintain an accuracy of one meter and lower in locating the target, even knowing nothing about the building structure. We also found that preventing the risks of illegal surveillance from this passive tracking without affecting the user communication experience remains challenging. We believe that addressing user privacy risks in massive MIMO-based technologies is an interesting and urgent matter, given their popularity in 5G and beyond. Then building an affordable defense scheme to protect users against illegal surveillance is a promising topic in the future.

REFERENCES

- [1] O. Kanhere and T. S. Rappaport, "Position location for futuristic cellular communications: 5G and beyond," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 70–75, 2021.
- [2] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.
- [3] J. Palacios, P. Casari, H. Assasa, and J. Widmer, "LEAP: Location estimation and predictive handover with consumer-grade mmwave devices," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 2377–2385, 2019.
- [4] S. Sadowski, P. Spachos, and K. N. Plataniotis, "Memoryless techniques and wireless technologies for indoor localization with the internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10996–11005, 2020.
- [5] C. Li, E. Tanghe, J. Fontaine, L. Martens, J. Romme, G. Singh, E. De Poorter, and W. Joseph, "Multistatic uwb radar-based passive human tracking using COTS devices," *IEEE Antennas and Wireless Propagation Letters*, vol. 21, no. 4, pp. 695–699, 2022.
- [6] B. H. F. Troels Pedersen, "Whitepaper on new localization methods for 5G wireless systems and the Internet-of-Things," *COST Action CA15104*, 2018.
- [7] L. Wielandner, E. Leitinger, F. Meyer, B. Teague, and K. Witrals, "Message passing-based cooperative localization with embedded particle flow," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5652–5656, 2022.
- [8] F. Hejazi, K. Vuckovic, and N. Rahnavard, "DyLoc: Dynamic localization for massive MIMO using predictive recurrent neural networks," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, pp. 1–9, 2021.
- [9] V.-L. Nguyen, P.-C. Lin, B.-C. Cheng, R.-H. Hwang, and Y.-D. Lin, "Security and privacy for 6g: A survey on prospective technologies and challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2384–2428, 2021.
- [10] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1038–1052, 2018.
- [11] P. Bahl and V. 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, pp. 775–784 vol.2, 2000.
- [12] William C. Stone, "National institute of standards and technology, construction automation program report no. 3, electromagnetic signal attenuation in construction materials," <https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir6055.pdf>, 1997.
- [13] N. Hosseini, M. Khatun, C. Guo, K. Du, O. Ozdemir, D. W. Matolak, I. Guvenc, and H. Mehrpouyan, "Attenuation of several common building materials: mmwave frequency bands 28, 73, and 91 GHz," *IEEE Antennas & Propagation Magazine*, vol. 63, no. 6, pp. 40–50, 2021.
- [14] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive mimo-ofdm system with deep convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10846–10857, 2019.
- [15] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, and A. Argyros, "A review on deep learning techniques for video prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2806–2826, 2022.
- [16] B. Jin, Y. Hu, Q. Tang, J. Niu, Z. Shi, Y. Han, and X. Li, "Exploring spatial-temporal multi-frequency analysis for high-fidelity and temporal-consistency video prediction," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Los Alamitos, CA, USA), pp. 4553–4562, IEEE Computer Society, jun 2020.
- [17] A. Alkhateeb, "Deep MIMO dataset," <https://deepmimo.net/>, accessed on November 3, 2022.
- [18] A. Guerra, F. Guidi, and D. Dardari, "Single-anchor localization and orientation performance limits using massive arrays: MIMO vs. Beamforming," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5241–5255, 2018.