Passive Device-Free Multi-Point CSI Localization and Its Obfuscation with Randomized Filtering

Marco Cominelli, Francesco Gringoli, Renato Lo Cigno
DII – University of Brescia, Italy

***Abstract*—The use of Channel State Information (CSI) as a means of sensing the environment through Wi-Fi communications, and in particular to locate the position of unaware people, is moving from feasibility studies to high precision applications. The work we present in this paper explores how the use of multiple localization receivers can enhance the precision and robustness of device-free CSI-based localization with a method based on a state-of-the-art Convolutional Neural Network. Next we discuss how a randomized pre-filtering at the transmitter can hide the information that the CSI carries on the location of one person indoor formalizing the manipulation technique. Results are presented discussing two different ways of exploiting the multi-receiver redundancy and how, in any case, properly randomized pre-distortion at the transmitter can prevent localization even if the attack is carried out with multiple localization devices (receivers controlled by the attacker).**

I. Introduction and State of the Art

Sensing as a side-service of Wi-Fi is becoming an industrial reality and in particular Channel State Information (CSI)-based localization is attracting attention for device-free indoor positioning. This research field was opened about ten years ago by seminal works like [1]–[4], then the academic community indulged on many variations of the topic, hinting to the possibility of identifying activities or gestures [5], [6], seeking health-care applications [7] and many others. The most recent trend is exploiting Machine Learning (ML) and Artificial Intelligence (AI) [8]–[11] to compensate the difficulty of finding analytic models with the power of supervised learning techniques for classification purposes, often involving Deep Learning or Reinforcement Learning. The recent survey [12] can compensate, for the interested reader, what we cannot discuss here for lack of space.

Two topics received instead little attention:

1) If and how multi-point reception can improve the reliability of CSI-based localization; and
2) How much localization impacts on the privacy of people.

Although the scarce attention to the first one may look surprising, we could find only two works on the topic. The authors of [13] propose to use massive Multiple Input Multiple Output (MIMO) technologies to improve the quality of CSI-based localization with Neural Network (NN). The work exploits up to 64 antennas, but indeed with a single logical measurement point; the work focuses on the learning technology and assumes that there is a service dedicated to localization, i.e., special frames are transmitted dedicated only to localization, thus this work should be compared mainly with technologies dedicated to localization as those based on time-of-flight like [14], rather than sensing as a side-effect of Wi-Fi communications. The research in [15] is indeed the only one that has similarities with our contribution, though it focuses on localization of a device, and not, as we do, on the localization of a person who does not carry a device. The work builds on the concept of channel charting [16], which lends to the possibility of semi-autonomous training because it uses differential positions and differential CSI, hence assuming a slowly changing channel with a CSI sampling that respect the Nyquist theorem, a condition that, for instance, cannot be assumed if Wi-Fi traffic is sporadic, or to detect the presence of a person in a room, a condition that implies a sort of discontinuity in the CSI.

Just as surprising is the overlooking of privacy implications, which has been addressed only very recently by our and other groups [17]–[23]. A first idea that comes to mind is using a reactive jamming device that selectively kills frames that belong to the localization attack, adapting for instance techniques like [17], [18]. To the best of our knowledge, this idea has never been explored in the literature, maybe because it kills traffic, thus if the frames used for localization also carry user data the communications will be heavily hampered. Additionally, such a technique requires to know that an attack is under way and the ability to identify the frames used for localization, otherwise it would become simply a jamming denial-of-service. A system to counter Wi-Fi sensing was originally proposed in [20] to prevent gesture recognition. Similar to [22], this system is based on an independent device that relays frames with the goal of superimposing an additional "reflection" of the signal that *obfuscates* the information imprinted by the environments on frames, differently from the more common *jamming* that superimpose a different signal, sometimes just noise, with the goal of killing the frame reception.

The works that lay the foundations of this paper are [19], [21], [22]. In the first two, we proposed for the first time a technique to obfuscate, or hide, the information carried by the CSI that enables localization, focusing on passive attacks, i.e., those where the attacker controls only the receiver. The core idea is to randomly distort the transmitted frames so that the CSI at the receiver looks like the one of a signal that has propagated through a different environment, unrelated to the one in which sensing is performed. In the third one, we tackled the problem of countering active attacks, i.e., those where the attacker controls both the transmitter and the localization receiver. In this case the countermeasures cannot be based on the pre-distortion of the transmitted frames, because transmitted frames are controlled by the attacker. The solution we proposed is based on a fast relay node, ideally an intelligent reflective surface like those discussed in [24], which introduces a time-varying additional reflection in the electromagnetic environment that prevents a localization device to pinpoint the position of a person.

Finally, exploiting techniques similar to those used in [21], [23] manipulates the CSI with the goal of avoiding device radiometric fingerprinting and preventing impersonation attacks. The topic of the paper is not localization, but if a person holds a Wi-Fi device a double attack identifying the device

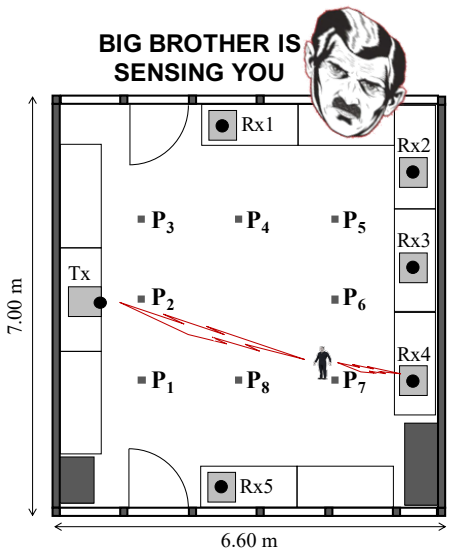

Figure 1: In the considered scenario the attacker can collect CSI data simultaneously at five receivers Rx1–Rx5 to locate a victim standing in one of eight possible target locations.

and the location of the person is more than a possibility.

The contributions of this paper in the context of localization and its obfuscation are two:

- First, we present the first experimental study that shows how, using multiple localization devices, the precision of localization can be improved;
- Second, we show that also in these conditions, the privacy of users can be preserved with a refinement of the CSI randomization technique we first introduced in [19], [21], not only protecting users from localization attacks, but also preserving the communication performance.

II. Attack Model and Scenario

In this section, we quickly summarize the reference attack model that can be used to monitor the activities and the position of unaware people through the opportunistic reuse of the Wi-Fi signals pervading modern environments. Then we describe how we implemented this model in our laboratory.

A. Attack Model

The attack model is shown in Fig. 1. We assume that the attacker (e.g., an employer whose goal is circumventing legislation on employees monitoring) can control multiple devices with the ability to extract CSI data from the received Wi-Fi frames. The large availability of extremely cheap and small platforms that can be converted into sensing nodes, like the Raspberry Pi [25], makes this feasible and cheap even on a large scale. A detailed analysis of the CSI structure at each single receivers, as well as a comparison between the CSI collected at different receivers, enables the attacker to determine the precise position of a person in the room. The attack considered in this paper extends the techniques described in [21]. The attacker collects a set of CSI traces with the help of a collaborator standing in specific target positions; then he trains a Convolutional Neural Network (CNN) with the collected data to use it later in order to determine the position of an unaware victim, e.g., an employee or collaborator.

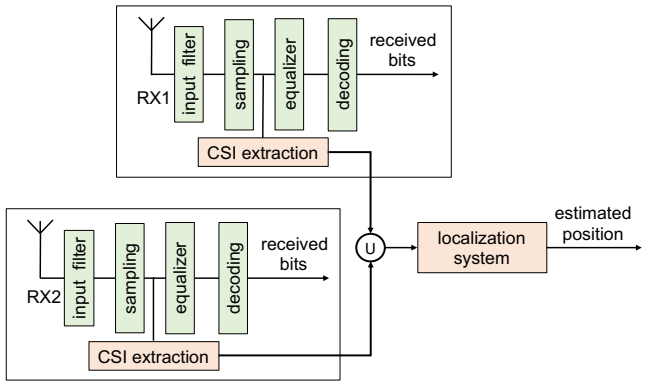

Figure 2: 802.11 modified receiver to infer people location: each receiver collects CSI and pushes everything to the localization system. Multiple CSI data originating from the same frame are "fused" to improve the accuracy with respect to a single receiver configuration.

B. Experimental Facility

We carry out the experiments in a laboratory of the ANS[1] group at the University of Brescia, whose map is shown in Fig. 1. Five receivers (from Rx1 to Rx5) are positioned on desks aligned along three of the four sides of the lab, while the transmitter (Tx) lies on a desk on the fourth side. We assume that the attacker controls all the receivers and knows their positions with respect to the transmitter, that can be one of the Access Points of a corporate network. The goal of the attacker is classifying the position of the person on 8 possible positions within the room (P1,...,P8). This configuration ensures that the person being tracked is always obstructing the line-of-sight (LoS) between the transmitter and at least one of the receivers; in this way the collected CSI should always be significantly affected at one or more receivers, independently of the victim's position.

III. CSI-based Localization and its Obfuscation

The principle behind the localization technique is the interaction between the Wi-Fi signals and the human body. In fact, the presence in an environment of a human body that absorbs, scatters, and reflects electromagnetic waves induces peculiar variations in the spectrum of the received signal that depend on the body position and movements. These variations can be studied by analyzing the CSI evaluated by every Wi-Fi device upon receiving a frame. As we show in Fig. 2, the correct decoding of frame's data requires an equalization of the spectrum of the received signal to reduce the distortion introduced by the communication channel. Extracting the CSI data from the chipset internals, one can directly observe the modifications that depend on the position of the observed person, given that the rest of the environment, including the position of the transmitter and the receiver, remains relatively stable over time.

The localization frameworks that have received more attention are based on NNs trained with CSI obtained when someone is standing in a known position; then, during the attack, the victim's position is estimated by recognizing the same patterns in the CSI [26]. It is important to capture a large amount of data not only to speed up the training phase, but also to obtain multiple CSI snapshots corresponding to the same position of the victim, as this would allow to "average out"

[1]The Advanced Networking Systems group is a research group in telecommunications at the University of Brescia, Italy.

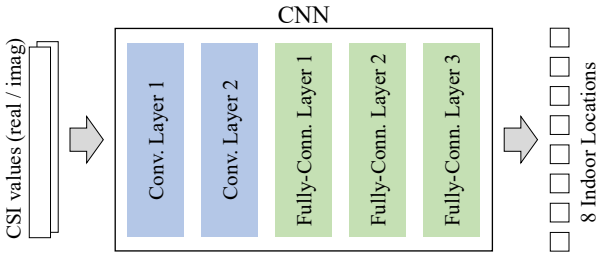

Figure 3: Architecture of the CNN used by our localization system.

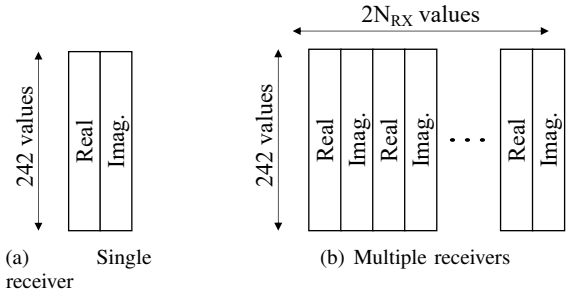

Figure 4: The network is trained with different types of input depending on the considered scenario; when we apply the CSI Data Fusion technique all the collected CSI are fed to a single NN.

minor spectral variations. This is in general not a problem in modern work places, where Wi-Fi signals pervade the environment and an attacker can opportunistically use the signals transmitted by the Access Points (APs) of a corporate network, given that such Wi-Fi nodes are usually in well-known positions and generate the largest amount of traffic.

In this work, we build the multi-receiver localization system on top of what we introduced in [21], where we have developed an efficient implementation for a single receiver based on a CNN with good localization accuracy[2]. A high-level representation of the NN architecture is shown in Fig. 3. We consider 802.11ac frames transmitted on 80 MHz channels with a single spatial stream encoded using Orthogonal Frequency Division Multiplexing (OFDM); each CSI data point is an array of 256 complex values, one per OFDM subcarrier. After removing the unused subcarriers (eleven at the edges of the spectrum and three at the center) we get as input for the CNN a 242×2 matrix. The first two convolutional layers of the CNN shown in Fig. 3 are used to extract complex features from the input data by exploiting the similarity of adjacent frequencies. In cascade to the convolutional layers, there are three fully-connected layers. The output of the last layer corresponds to a choice among one of the possible classes, i.e. positions that are decided during the training phase. All the layers but the last one (which uses a softmax function) use a common Rectified Linear Unit (ReLU) activation function. Finally, we use the Adaptive Momentum Estimation (ADAM) algorithm to adjust the weights of the CNN during the training phase.

The considered CNN achieves a good accuracy and this is clearly related to the unique and remarkably constant CSI data that are obtained for each position of the person under tracking in the room. Simple reasoning suggests that a random pre-distortion of the transmitted signals should suffice to disrupt localization accuracy and obfuscating the person's position. In the feasibility study presented in [19], [21] we obtained excellent obfuscation results by placing peaks at random positions in the spectrum of transmitted frames, following a simple time correlation structure. As a consequence of this coarse manipulation, the presence of the obfuscator could have been detected (and eventually countered) with a careful analysis of the signal. Furthermore, the communication performance when the obfuscator was active was severely degraded, leading to highly reduced throughput.

IV. Improving Localization with Multiple Receivers

The literature on CSI-based localization has so far considered only systems with a single receiver, but combining CSI data collected at multiple receivers seems a promising extension to improve the accuracy. With enough receivers it is possible, in fact, to always have at least one receiver whose LoS to the transmitter intercepts the person under tracking. For instance, if the target person stands in position P_7 in Fig. 1, we expect minor interference effects on $Rx1$, but clear effects on $Rx4$ due to the obstructed LoS. One of the goals of this work is to extend the localization system by combining the CSI captured at multiple receivers as shown in Fig. 2; the second one is showing that also this powerful attack can be countered.

We present two methods for "combining" the localization data: we discuss here their pros and cons while we present the experimental results in Sect. VI. Common to the two methods is that receivers are positioned in the same indoor environment and they have the capability of matching the reception of the same packets (we use timestamps, but other techniques are just as fine).

A. Majority Vote

In this implementation we combine the output of multiple stand-alone localization systems, i.e., one receiver is associated to one CNN. Given that receivers in the room are located several wavelengths apart from each other, we assume that the CSI vector at the input of each localization system, which we report in Fig. 4(a) as 242×2 matrix, is independent from the others. Thus the CNNs actually learn and classify independent models, so that errors are independent, and we can assume that "summing" the results compensates the random errors. Unfortunately, positions cannot be "summed" in an algebraic way, but we can decide based on the majority of decisions. Given that position errors are independent by construction, a majority vote also corresponds to a Maximum Likelihood Estimation (MLE) and should be optimal given the assumptions.

Let N_r be the number of receivers. Independently from N_r, it is always possible that the vote does not have a majority, e.g., 3 receivers have classified 3 different locations, and no decision can be taken. We separate wrong classifications (a decision is taken, but it is for a wrong position), from undecidable situations when no classification is possible.

B. CSI Data Fusion

A more sophisticated use of the information at different receivers is based on the fusion of the CSI vectors, assuming that an extended CNN can do a better job than simple majority voting. Note that majority vote is a MLE, but only under the assumptions of independent decisions. The data fusion

[2]Further details on this project, the software produced an so forth can be found at https://ans.unibs.it/projects/csi-murder/.

changes the knowledge base of the estimator, thus we can hope in a more powerful technique.

In this second implementation we consider an extended CNN that processes the CSI vectors collected at the N_r receivers as a "fused" and larger dataset, i.e., a $N_r \times 242 \times 2$ matrix as we show in Fig. 4(b). Differently than the first implementation, this method always takes a decision for a specific position, as in the case of a single receiver, and does not have undecidable situations. We have not attempted to design a new CNN for this task, so, whatever the results we obtain, we cannot exclude that a different learning method, based on CNNs or on a different technique, can obtain better results.

V. Principles of CSI Randomization and Its Implementation

While any approach that can improve the localization accuracy represents a positive result, at the same time it can be considered an increasing hazard against the privacy of the tracked people. In addition, an improved localization technique can also have detrimental effects against simple obfuscation techniques like the one that we introduced in [21]. The goal of this section is hence to design an improved obfuscation technique that can be effective independently of the number of involved receivers. Here we study how to defeat the two CNN based localization mechanisms introduced in Sect. IV to restore the privacy of the tracked person. A good obfuscation technique is as unobtrusive as possible, effective in preventing localization, and also maintains good communication performance.

As discussed in [21], the frequency dependent amplitude of the received signal is the feature that is mostly considered by the CNN in classifying CSI and mapping the person location. For this reason, we focus the randomization technique on the amplitude of the subcarriers that compose the spectrum of the transmitted signals. We also consider a single transmission chain for the sake of simplicity: the extension to multiple transmission chains is left for future work.

Let N_{sc} be the number of carriers used by the Wi-Fi OFDM modulation, $f_i,\ i = 1, 2, \ldots, N_{\text{sc}}$ the carrier number; $k = 1, 2, \ldots$ the discrete time index identifying the frame and $\Delta_t(k)$ the absolute (continuous) time between frame k, and frame $k-1$; $\Delta_t(1)$ is undefined, but it is irrelevant for our purposes. We only consider carriers that are not suppressed by the system, thus excluding the middle carrier and those in the guard bands. The magnitude of the spectrum at the receiver, or CSI, derived from the known initial symbols of the frame, represents the frequency and time varying signal attenuation (or channel response) introduced by the channel $A_R(f_i, k)$. Notice that the entire Wi-Fi PHY layer is based on the assumption that the channel coherence is long enough to guarantee that the channel response is constant during a single frame, thus our modeling does not introduce significant approximations.

The goal of the obfuscator is to guarantee that the information in $A_R(f_i, k)$ does not allow an attacker to properly classify the position of a person in the room. Ideally, the obfuscator should guarantee that the mutual information between the CSI and the location of a person in the room is zero, but this theoretical analysis is out of the scope of this paper.

To achieve this goal we multiply the IQ samples of the digital signal before performing the Inverse Fast Fourier Transform (IFFT) conversion on the OFDM symbol, so that the actual CSI information at the receiver is:

$$A_R(f_i, k) = A_C(f_i, k) \times A_O(f_i, k) \tag{1}$$

where $\times$ is the standard algebraic multiplication applied separately carrier by carrier: $A_O(f_i, k)$ is a pre-distortion mask whose goal is adding random information to $A_C(f_i, k)$ so that $A_R(f_i, k)$ maintains the properties that allow demodulation and correct decoding of the frame given the CSI, but information on the real channel response is degraded to a point where localization is not better than a random guess.

The pre-distortion $A_O(f_i, k)$ must have the following characteristics:

1) It does not alter the frame power:
$$\sum_{i=1}^{N_{\text{sc}}} A_O(f_i, k) = K_A \tag{2}$$
meaning that if some frequencies are amplified, then others must be attenuated, with K_A some appropriate constant.
2) It guarantees that the correlation in time is compatible with the standard movement of a person in a room;
3) It guarantees that the attenuation in frequency is compatible with the channel Doppler spread;
4) It cannot be inverted within a reasonable time, i.e., given the sequence $A_R(f_i, k); k = h, \ldots h + H$ it should not be possible to reconstruct the sequence $A_O(f_i, k)$, not even when using multiple receivers; H is a design parameter whose impact on the system is left for future study;
5) It does not modify the communication performance of the system.

The formalization and (if possible) the proof that the five conditions above are feasible are beyond the scope of this work. In the following we present a heuristic Markovian methodology that we evaluate in Sect. VI.

Let $\mathbf{R}$ be a vector of independent, continuous random variables ρ_i of dimension N_{sc}, one for every OFDM subcarrier. Each random variable is drawn from the same distribution $f_R(\rho)$, and they are all independent one another. For reasons that will be clear shortly, we select $f_R(\rho)$ to be a uniform distribution with support $(\rho_{\min}, \rho_{\max})$. Consider now the multidimensional random process defined as:

$$\mathcal{R}(k) = e^{-\alpha \Delta_t(k)} \mathcal{R}(k-1) + \mathbf{R} \tag{3}$$

Equation (3) defines a Uniform-Markov process, which, compared with the more popular Gauss-Markov process[3] exhibits uniform increments taken from $f_R(\rho)$ instead of Gaussian increments. The dimension of the process is N_{sc}. The process is discrete time, because the transmission of frames defines a discrete time index; however, the process memory depends on the absolute time $\Delta_t(k)$ in order to guarantee that frames transmitted far apart in time do not have excessive correlation. As it is well known from probability theory, this process exhibits a dependence in time that increases when α decreases, which allows tuning the obfuscation to the expected movements of people in the room.

[3]For instance, the error of Global Navigation Satellite System positioning is normally modeled with a three-dimensional Gauss-Markov process.

The random process defined by Eq. (3) has no correlation in the frequency domain, which is in contrast with the desired characteristic 3) defined above. To overcome this unwanted feature, we use a simple convolutional filter as follows:

$$A_O(k) = [1 + \mathcal{R}(k)] * \Theta_C \tag{4}$$

where $*$ is the standard convolutional sum and Θ_C is a symmetric filter with length C; C must be odd for symmetry and $3 \leq C \leq N_{sc}$. The dependency on f_i is implicit in the convolution, and appropriate leading and trailing zeros must be appended to $\mathcal{R}(k)$ to allow the convolution. The shape and characteristics of Θ_C can be studied to optimize the performance. In this work we do a simple moving average (all coefficients are 1) with $C = 5$.

Equations (3) and (4), together with the normalization (2) (to be run at every step) define, in our opinion, an appropriate randomized localization obfuscation that respects the characteristics illustrated above. $\rho_{\min}$, $\rho_{\max}$, α, C, and the values of Θ_C taps can be used for tuning the system, and can also be changed over time to make it harder for an attacker to invert the obfuscation. A proper sensitivity analysis on all these parameters is not feasible within a single paper, and we are more interested in fundamental properties than in finding the optimal setting, which may also depend on the considered scenario. A quick preliminary study was sufficient to select $\rho_{\min} = -0.3$, $\rho_{\max} = 0.3$, and $\alpha = 0.2$ for achieving acceptable performance. Notice that $\alpha = 0.2$ means that if $\Delta_t(k) \geq 15\,\text{s}$, then $A_O\ (k)$ and $A_O\ (k\text{-}1)$ are almost completely uncorrelated (the correlation coefficient is below 5%, which is coherent with requirement 2) above.

Equations (3) and (4) cannot guarantee that the pre-distortion does not lead to amplitudes smaller than zero, which is obviously not implementable. Also zero or very small values are not desirable, as they imply that an entire subcarrier is suppressed, which introduce systematic transmission errors that should be avoided. To prevent this possibility we use a simple clipping function $[\cdot]_{\min}^{\max}$, that cuts the amplitude of each subcarrier between a $\min$ and a $\max$ value. Ideally, only the $\min$ clip is necessary to avoid negative and too small values, but this asymmetry makes it difficult to guarantee that the average power of frames is not altered. Furthermore, the clipping can be applied before or after the convolutional filtering of Eq. (4). Predicting the consequences of clipping before or after the filter is very difficult, because clipping is highly non linear, thus we decided to explore the performance of both options implementing the equivalent of the following two equations:

$$A'_O(k) = [1 + \mathcal{R}(k)]_{\min}^{\max} * \Theta_C \tag{5}$$

$$A''_O(k) = [1 + \mathcal{R}(k) * \Theta_C]_{\min}^{\max} \tag{6}$$

As for the clipping values, we explore two possibilities: a symmetric case with $\min = 0.1$ and $\max = 1.9$, and an asymmetric case with $\min = 0.1$ and $\max = \infty$, where we only guarantee that the maximum power density over the frame spectrum does not exceed the values achieved by non-obfuscated frames, but we cannot guarantee that the average frame power is maintained.

A. Implementation

We describe here some implementation details, focusing on the constraints that may approximate the formulas introduced above. First of all we note that $N_{sc} = 256$ as we work with $80\,\text{MHz}$ 802.11ac frames. Second, we implemented the receivers using COTS Access Points from Asus: we chose the rt-ac86u model as it can extract CSI data from the transmitted frames. Third, we implemented the transmitter with an Ettus USRP N300 Software-Defined Radio (SDR) whose bandwidth exceeds the $80\,\text{MHz}$ requirement. We chose an SDR because we need precise control over the generation of each Wi-Fi frame in order to craft and apply the pre-distortion that modifies the CSI. We generate Wi-Fi frames with the MATLAB WLAN Toolbox running on a workstation equipped with an Intel Core i7 and 16 GB of RAM. We also prepare obfuscated frames directly in MATLAB before sending the corresponding sequence of IQ samples to the SDR. As the SDR does not run a MAC algorithm, there could be some uncontrolled collisions on the channel, even if we have selected a channel (157) that is not used in our University.

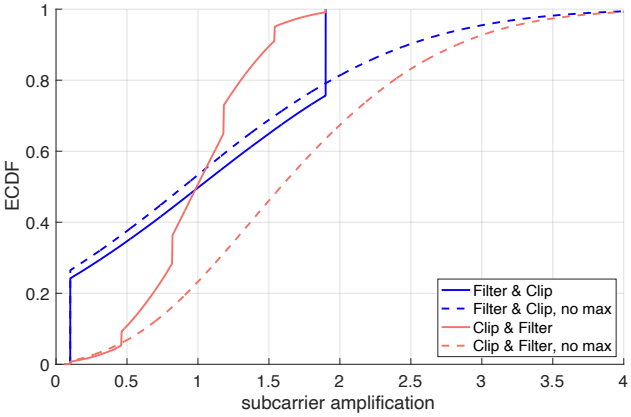

Figure 5: ECDF of the amplification factors assigned to the OFDM subcarriers in Eq. (5) and (6) with both symmetric and asymmetric clipping. Blue lines refer to Eq. (6) and red ones to Eq. (5); 'no max' means that $\max = \infty$.

The MATLAB code implements the obfuscation techniques described by Eq. (5) and (6) by applying the pre-distortion, if present, before the IFFT—that is, in the frequency domain. We avoid working in the time domain because it would require the usage of a circular convolution. Before sending the stream of IQ samples to the SDR, we normalize them to the highest value, i.e., we divide them by the one with the largest absolute module, so that we use the entire range of the SDR.

We highlight that the overall procedure cannot strictly guarantee that the frame power is not altered as required by Eq. (2): this entails evaluating the power spectrum of the entire frame, which we are unable to do; however, we deem that the pre-distortion, especially with symmetric clipping, does not change significantly the power spectrum on the channel. Fig. 5 shows the ECDF of the marginal distribution of the amplitude of the processes described by Eq. (5) (Clip & Filter, in red) and Eq. (6) (Filter & Clip, in blue). The ECDF is computed over 10,000 frames, or equivalently, $256 \times 10,000 = 2,560,000$ samples. The solid lines refer to symmetric clipping: their median value is one, as expected. As the OFDM modulation is not constant envelope and its peak to average ratio depends on the payload, and can be as large as $12\,\text{dB}$, a pre-distortion that preserves the average amplitude should not alter significantly the power spectrum after the modulation, but this aspect requires further investigation. The dashed lines (no max) refer

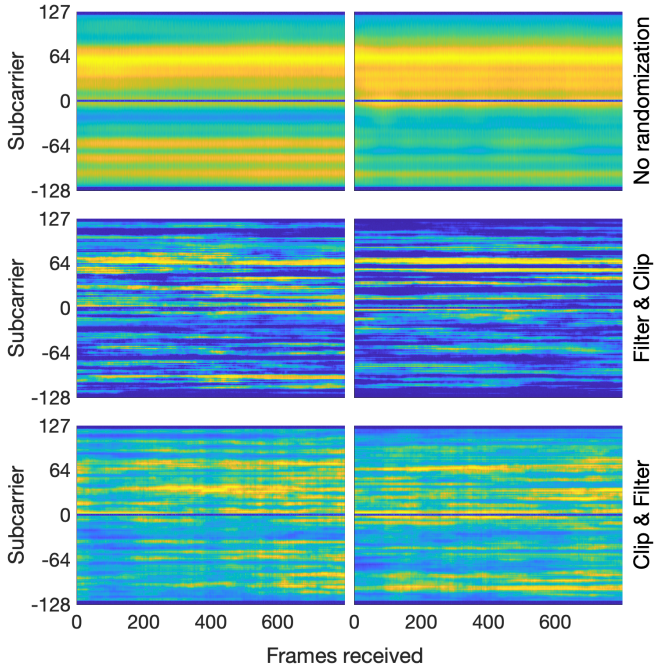

Figure 6: Magnitude of the CSI collected from 800 frames with a person standing in positions P1 and P2 (brighter yellow-ish colors mean larger magnitude). The two plots at the top refer to clean transmissions: the horizontal bands show that CSI are constant over time if the person does not move enabling localization by ML. The two plots at the bottom refer to the same positions when the obfuscation is on, the middle one with a Filter & Clip and the bottom one with Clip & Filter.

instead to the case when clipping is applied only to the min amplitude, and the median is obviously not one, but we do not know if this alters significantly the frame power spectrum. The counter-intuitive behavior of the solid red curve (Clip & Filter) deserves a final comment. Indeed, by first clipping and then filtering, we introduce multiple discrete components (and not just two as in Filter & Clip) in the distribution: they are given by the combinations of clipped subcarriers in the 5-tap average filter, and are reflected in the ECDF.

Fig. 6 reports qualitative results measured at the receiver. It visually shows why Wi-Fi sensing can locate people, and why the proposed obfuscation technique is a valid countermeasure. Without obfuscation the channel response (amplitude of the CSI) is remarkably constant over time (x-axis). The CNN can learn and classify the position of the person. Obfuscation, instead, keeps changing the CSI amplitude, thus preventing proper learning and classification. We would like to recall that all the software we have used and the data we have collected are available on our website https://ans.unibs.it.

VI. Experimental Results

Table I presents the experimental localization accuracy. Overall, we collected $5 \times 8 \times 800 = 32,000$ frames for training the system, and $5 \times 8 \times 100 = 4,000$ for testing. Training and testing are done separately both for the clean case and for each obfuscation technique (5 different cases in total) and separated by a reasonable time (a few tens of minutes at least) so that training and testing correspond to a realistic attack, and the risk that localization results are obtained just by chance, because of time correlation between training and testing is minimized. We merge the data collected in different positions by post-processing the acquired traces, thus reducing the number of experiments to be performed. Also, this allows applying the two different multi-point localization techniques (see Sect. IV-A and IV-B) exactly on the same measured data, thus avoiding that the difference observed in the two techniques is related to differences in the experimental environment and not an intrinsic property. Since we consider 8 possible positions, a random guess would lead to an average accuracy of 12.5%, and this is our reference for evaluating the quality of the obfuscation.

Table I: Localization performance comparison using five different receivers and different position merging techniques; C&F = Clip & Filter, F&C = Filter & Clip, nm = no max clipping. For the Majority Vote the results in parentheses is the percentage of frames without a final decision (no majority reached).

Localization Accuracy [%]

Single Receivers	Rx1	Rx2	Rx3	Rx4	Rx5
Clean	90.6	89.6	93.1	83.1	67.6
C&F	17.6	41.9	15.4	33.6	15.6
C&F, nm	12.9	57.4	30.6	60.5	21.4
F&C	8.8	28.0	7.4	15.2	0.0
F&C, nm	24.5	37.8	22.1	44.9	22.6

Majority Vote	$N_r = 2$	$N_r = 3$	$N_r = 4$	$N_r = 5$
Clean	71.1 (28.9)	95.7 (4.3)	99.9 (0.1)	100.0 (0.0)
C&F	8.7 (72.5)	18.1 (39.1)	19.2 (22.9)	18.2 (20.1)
C&F, nm	13.7 (75.0)	29.8 (42.0)	36.4 (24.4)	34.0 (23.0)
F&C	1.1 (70.2)	3.1 (35.9)	2.6 (23.2)	1.5 (31.2)
F&C, nm	18.2 (60.6)	25.4 (33.1)	27.3 (23.9)	31.4 (16.5)

Data Fusion	$N_r = 2$	$N_r = 3$	$N_r = 4$	$N_r = 5$
Clean	83.0	84.4	83.8	91.1
C&F	27.1	26.3	30.7	8.0
C&F, nm	30.3	27.1	26.6	4.4
F&C	7.0	4.5	1.4	1.1
F&C, nm	23.8	27.9	25.2	22.1

Analyzing the data for each single receiver highlights (the top sub-table in Tab. I, as we already observed in [22], that the position of the receiver influences the localization performance: Rx5 performs much worse than the others, while Rx3 significantly better. Furthermore, we observe that the localization accuracy decreases with every obfuscation technique regardless of the position of the receiver and that symmetric clipping offers superior performance, although we do not have a formal explanation for this latter fact. While a thorough explanation of this behavior deserves additional research, we speculate it is the same phenomenon that makes Filter & Clip obfuscator working better than Clip & Filter. Actually, the obfuscation achieved by Filter & Clip is almost perfect, with only Rx2 scoring a better accuracy than a random guess (28%), yet hardly usable for any meaningful insight on the real location of the person. We will see later that this comes at the expense of the achievable throughput, and we deem that both results are related to the discontinuity introduced in the spectrum when clipping is performed as last operation, which means that some high frequency harmonic disturbs the signal, although this cannot be easily observed on frame-wise analysis.

Coming to the results achieved with the multi-receiver attack, the first thing that emerges is that a simple and traditional Majority Voting outperforms the apparently more sophisticated Data Fusion approach, which does not seem to improve the performance even in absence of obfuscation. As

we already commented in Sect. IV, this does not exclude that some other data fusion technique may outperform Majority Voting, but just that improving CNN-based localization simply adding information may be more difficult than expected. A possible explanation lies in the fact that the Data Fusion approach does not make the localization errors of different receivers independent one another, thus the strong correlation in all the data induced by the CNN makes some errors dominant.

The results in parentheses for Majority Voting in Tab. I are the percentage of un-decidable situations, i.e., a majority decision has not been reached. As expected, these are particularly high when we consider only two receivers, while increasing the number of receivers increases the performance and decreases the number of un-decidable cases, until localization without obfuscation becomes "perfect" for $N_r = 5$ receivers, and we can consider the localization deterministic also with 4 receivers. Notice that given the amount of data we collected, the absence of even a single mis-classification makes the result for $N_r = 5$ extremely reliable and significant, making Wi-Fi-based localization a true threat for location privacy. For $N_r \leq 5$ the results reported are the average over all possible combinations of the five receivers in pairs, triplets and quadruplets, again making the results highly reliable, and indicating that multi-point localization also makes the technique less sensitive to the receiver position.

Considering the obfuscation results, it is immediately clear that the technique we propose is robust also against a sophisticated attack brought with five receivers strategically placed in the surveyed room. Again, the Filter & Clip technique completely obfuscates the localization, indeed making it even worse than a random guess[4] and outperforming any other strategy. Note that when the obfuscation is active, for Majority Voting, there is a high percentage of wrong decisions (the sum of the two percentages reported is smaller than 100%), which were instead completely absent without obfuscation. This is a strong indication that the randomized pre-distortion implemented by the obfuscator successfully deceives the localizer, so that it does not learn anything really meaningful during the training phase, in spite of the fact that the obfuscator is on, thus the classifier (recall that the CNN learning is supervised, so training should be effective in any case if there is information to exploit) indeed learns random patterns and it is not able to single-out the channel characteristics.

A. *Impact on throughput*

In Wi-Fi, the achievable throughput depends on the chosen Modulation and Coding Scheme (MCS). The 802.11ac standard defines ten MCS over 80 MHz and 800 ns guard period with corresponding throughput increasing from 29.3 Mbit/s (MCS 0) to 390 Mbit/s (MCS 9). A higher MCS enables higher throughput, but it is more sensitive to distortion, noise and interference because of more advanced modulation techniques and less robust correction codes.

We show in Fig. 7 the impact of different randomization techniques on the Packet Delivery Rate (PDR), i.e., the percentage of Wi-Fi frames correctly received. We transmit

[4]We do not consider this result really positive, because it can be an indication that there is still some information about the location in the CSI, but the localization system is unable to use it; yet we think that it will be very hard for an attacker to find out how to exploit this location information.

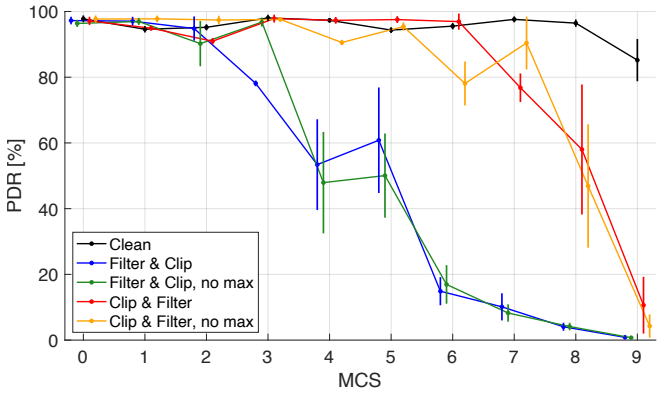

Figure 7: PDR as a function of the MCS with different types of randomization applied to the transmitted signal. Each point represent the average result, while the vertical bars identify the 90% confidence interval. Points are slightly offset from the integers they refer to for the sake of readability

1000 frames for every possible MCS and for every randomization technique we are considering, and we count how many frames we correctly decode at each receiver. The results show that the choice of the pre-distortion technique influences the communication performance, especially for higher MCS. There are two interesting details to notice here.

First of all, the PDR performance does not change significantly whether the randomization masks have their components clipped or not to a maximum value, thus selecting symmetric clipping is desirable as it yields better localization obfuscation while not affecting communication performance. Indeed, the 90% confidence intervals obtained with the same method overlap almost always for given MCS, indicating the minor variations in the average is due to randomness. Second, we observe a profound difference in the performance achieved by Clip & Filter versus Filter & Clip. It is evident that filtering the randomization mask before applying clipping has a destructive effect on the communication performance, and this is particularly evident for MCS ≥ 3, when QAM modulations are used, which are more sensitive to amplitude distortion.

While it is difficult to provide a solid justification for this evidence, we can try to explain this by noticing that, before clipping, the values assumed by the random process may be well-above or below the clipping threshold. The clipping operation forces the mask into the admissible range of values, but also introduce some high frequency components in the signal spectrum, which may interfere with other subcarriers, introducing systematic errors. In the transition phase between good and bad performance (MCS 3, 4, and 5), the behavior is particularly chaotic due to the interplay between the systematic errors and the convolutional codes that change rate with different MCSs. On the other hand, filtering the mask when the clipping has been already applied has the effect of producing a "smoother" mask with less high-frequency components; however, this also helps the localization that in some cases has an accuracy fairly better than a random choice.

VII. Conclusions and Future Work

Recent works have shown that:

1) It is feasible to exploit communication signals opportunistically to sense indoor environments and localize unaware victims;

2) CSI randomization represents an effective countermeasure against such types of passive (CSI-based) localization attacks.

In this work we make one step further by considering a scenario in which the attacker can combine information collected at multiple receivers to improve the localization accuracy. Our experimental results show that proper CSI randomization techniques can still disrupt localization attacks carried out with more than one receiver: Even when five receivers are used, a case that make localization without obfuscation practically perfect, proper randomization completely destroys the possibility of localizing a victim. This is achieved without disrupting communications, although the randomization that better obfuscate localization makes it also difficult to decode frames with high MCSs.

This work also lays the foundation for more theoretically-sound randomization techniques that are virtually identical to real channel responses instead of altering the signal with few prominent features that can be detected (and overcome) by an attacker. We emphasize that the requirements for the obfuscator that we discussed in this paper makes it is difficult to formalize properly the randomization process, because of the frequency-time correlation structure of the process itself. We leave this formalization and the optimization of the proposed randomization techniques as future work and as a challenge for the research community.

REFERENCES

[1] K. Chetty, G. Smith, and K. Woodbridge, "Through-the-Wall Sensing of Personnel Using Passive Bistatic WiFi Radar at Standoff Distances," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1218–1226, Apr. 2012.

[2] F. Adib and D. Katabi, "See through walls with WiFi!" In *ACM Int. Conf. of the Special Interest Group on Data Communication (SIGCOMM)*, Hong Kong, Aug. 2013, pp. 75–86.

[3] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor Localization via Channel Response," *ACM Computing Surveys*, vol. 46, no. 2, pp. 25:1–32, Dec. 2013.

[4] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. Ni, "CSI-Based Indoor Localization," *Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.

[5] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *IEEE Conf. on Computer Communications (INFOCOM)*, Hong Kong, Apr. 2015, pp. 1472–1480.

[6] F. Zhang, C. Chen, B. Wang, and K. J. R. Liu, "WiSpeed: A Statistical Electromagnetic Approach for Device-Free Indoor Speed Estimation," *IEEE Internet of Things Jou.*, vol. 5, no. 3, pp. 2163–2177, Jun. 2018.

[7] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," *IEEE Trans. on Mobile Computing*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[8] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based Fingerprinting for Indoor Localization: A Deep Learning Approach," *Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[9] G.-S. Wu and P.-H. Tseng, "A Deep Neural Network-Based Indoor Positioning Method using Channel State Information," in *Int. Conf. on Computing, Networking and Communications (ICNC)*, Maui, HI, USA, Mar. 2018, pp. 290–294.

[10] E. Schmidt, D. Inupakutika, R. Mundlamuri, and D. Akopian, "SDR-Fi: Deep-Learning-Based Indoor Positioning via Software-Defined Radio," *IEEE Access*, vol. 7, pp. 145 784–145 797, Oct. 2019.

[11] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning," in *IEEE Int. Conf. on Pervasive Computing and Communications (PerCom)*, Kyoto, Japan, Mar. 2019, pp. 1–10.

[12] Ma, Y. and Zhou, G. and S. Wang, S., "WiFi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 46:1–36, Jun. 2019.

[13] M. Widmaier, M. Arnold, S. Dorner, S. Cammerer, and S. ten Brink, "Towards Practical Indoor Positioning Based on Massive MIMO Systems," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, Nov. 2019, pp. 1–6.

[14] Ricciato, Fabio and Sciancalepore, Savio and Gringoli, Francesco and Facchi, Nicolò and Boggia, Gennaro, "Position and Velocity Estimation of a Non-Cooperative Source From Asynchronous Packet Arrival Time Measurement," *IEEE Trans. on Mobile Computing*, vol. 17, no. 8, pp. 2166–2179, Sep. 2018.

[15] E. Gönültaş, E. Lei, J. Langerman, H. Huang, and C. Studer, *CSI-Based Multi-Antenna and Multi-Point Indoor Positioning Using Probability Fusion*, 2020. arXiv: 2009.02798.

[16] C. Studer, S. Medjkouh, E. Gönültaş, T. Goldstein, and O. Tirkkonen, "Channel Charting: Locating Users Within the Radio Environment Using Channel State Information," *IEEE Access*, vol. 6, pp. 47 682–47 698, Aug. 2018.

[17] D. Nguyen, C. Sahin, B. Shishkin, N. Kandasamy, and K. R. Dandekar, "A Real-Time and Protocol-Aware Reactive Jamming Framework Built on Software-Defined Radios," in *ACM Workshop on Software Radio Implementation Forum*, Chicago, Illinois, USA, 2014, pp. 15–22.

[18] M. Schulz, F. Gringoli, D. Steinmetzer, M. Koch, and M. Hollick, "Massive Reactive Smartphone-Based Jamming Using Arbitrary Waveforms and Adaptive Power Control," in *10th ACM Conf. on Security and Privacy in Wireless and Mobile Networks (WiSec)*, Boston, MS, USA, 2017, pp. 111–121.

[19] M. Cominelli, F. Kosterhon, F. Gringoli, R. Lo Cigno, and A. Asadi, "An Experimental Study of CSI Management to Preserve Location Privacy," in *14th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization (WiNTECH)*, London, UK, Sep. 2020, pp. 1–8.

[20] Y. Qiao, O. Zhang, W. Zhou, K. Srinivasan, and A. Arora, "PhyCloak: Obfuscating Sensing from Communication Signals," in *13th USENIX Conf. on Networked Systems Design and Implementation (NSDI'16)*, Santa Clara, CA, USA, Mar. 2016, pp. 685–699.

[21] M. Cominelli, F. Kosterhon, F. Gringoli, R. Lo Cigno, and A. Asadi, "IEEE 802.11 CSI randomization to preserve location privacy: An empirical evaluation in different scenarios," *Elsevier Computer Networks*, vol. 191, no. 22, p. 107 970, May 2021.

[22] M. Cominelli, F. Gringoli, and R. Lo Cigno, "Non Intrusive Wi-Fi CSI Obfuscation Against Active Localization Attacks," in *16th IFIP/IEEE Conf. on Wireless On-demand Network systems and Services (WONS)*, Virtual Conference, Mar. 2021, pp. 1–8.

[23] Abanto-Leon, Luis F. and Bäuml, Andreas and Sim, Gek Hong (Allyson) and Hollick, Matthias and Asadi, Arash, "Stay Connected, Leave no Trace: Enhancing Security and Privacy in WiFi via Obfuscating Radiometric Fingerprints," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 4, 44:1–44:31, 3, Article 44 Dec. 2020.

[24] M. Di Renzo, M. Debbah, D. Phan-Huy, and et al., "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *J Wireless Com Network*, vol. 129, 2019.

[25] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *13th ACM Int. Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization (WiNTECH '19)*, Los Cabos, Mexico, Oct. 2019, pp. 21–28.

[26] C. Cai, L. Deng, M. Zheng, and S. Li, "PILC: Passive Indoor Localization Based on Convolutional Neural Networks," in *IEEE Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, Wuhan, China, Mar. 2018, pp. 1–6.