An Imperceptible Eavesdropping Attack on WiFi Sensing Systems

Li Lu, *Member, IEEE*, Meng Chen, *Graduate Student Member, IEEE*, Jiadi Yu, *Senior Member, IEEE*, Zhongjie Ba, Feng Lin, *Senior Member, IEEE*, Jinsong Han, *Senior Member, IEEE, ACM*, Yanmin Zhu, *Senior Member, IEEE*, and Kui Ren, *Fellow, IEEE, ACM*

***Abstract*— Recent years have witnessed enormous research efforts on WiFi sensing to enable intelligent services of Internet of Things. However, due to the omni-directional broadcasting manner of WiFi signals, the activity semantic underlying the signals can be leaked to adversaries for surveillance, as demonstrated by our previous work. In this paper, we further extend the attack capability of *ActListener* to impersonation attack, which could eavesdrop on users' behavioral uniqueness imperceptibly using a WiFi infrastructure in any location of user sensing area. In particular, *ActListener detects each human activity* and converts the eavesdropped signals to that by legitimate devices based on our proposed signal propagation models. To extract noise-resilient individual behavioral uniqueness from converted CSI of WiFi signals, we further add user identification models into the substitute model set for training the signal pattern calibration generative model. Experimental results demonstrate that *ActListener* could achieve over 80% accuracy in activity semantics retrieval and impersonation by using the converted signals.**

***Index Terms*— Imperceptible eavesdropping, activity recognition, user identification, WiFi signal.**

I. Introduction

With the rapid development, Internet of Things (IoT) has facilitated the intelligentization of various conventional appliances (such as television, speaker, etc.), turning them into smart IoT devices, with the assistance of prevalently deployed WiFi for network connection. Building on the basis of these pre-deployed infrastructures, enormous research efforts have been put into exploring WiFi for non-intrusive and low-cost sensing in IoT environments [1]. As its significant growth, the number of household WiFi-connected devices has grown rapidly. As investigated in a recent report [2], the average US home has a full dozen connected devices, and the number is expected to be 20 by 2025. And active investments are also attracted by WiFi sensing to support the development of related enterprises that develop artificial intelligence-based automation solutions for residential buildings, such as Perspicace Intelligent Technology [3]. All of these indicate the bright future of WiFi sensing.

Following such a trend, many researches make efforts on realizing WiFi-based activity recognition and user identification for natural human-computer interaction and non-intrusive privacy protection. Early work [4] demonstrates the feasibility of using WiFi to recognize human daily activities, inspired by which, following works [5], [6], [7], [8], [9], [10] explore the sensing capability of WiFi on more fine-grained activity and gesture recognition. Such systems not only enable users a more natural interactive experience, but also explore users' personal information by long-term tracing for personalized and intelligent services [11], e.g., a user's daily gestures interacted with a specific appliance can be employed to expose his/her gender, age, and even work efficiency. On the other hand, recent research efforts [12], [13], [14], [15], [16] further push WiFi sensing to individual identification area for rigorous access control, extending the usage of WiFi sensing. Both the coarse-grained activities [16], [17], [18] and fine-grained finger gestures [12], [13], [14], [15] are explored to realize WiFi-based user identification. These enable users to conduct sensitive operations (such as accessing the privacy system built on TEE, i.e., Trusted Execution Environment, or online shopping) in a non-intrusive manner. With sustained efforts and achievements, IEEE has even launched the IEEE 802.11 WLAN Sensing Study Group [19], and initiated to add WiFi sensing as a basic technique in the coming IEEE 802.11 series standard (i.e., 802.11bf) [20].

However, the security concerns of these connected devices gradually become the shadow behind their prosperity. Our previous work [21] has revealed the WiFi's omni-directional broadcasting manner introduces the inborn vulnerability to these systems. The omni-directional broadcasting WiFi signals could be sniffed by any device around the wireless

Manuscript received 15 January 2023; revised 1 August 2023; accepted 12 May 2024; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor T. Qiu. Date of publication 28 May 2024; date of current version 17 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62102354, Grant 62032021, Grant 62172359, Grant 62372406, Grant 62172277, Grant 62372400, and Grant 62072304; in part by the National Key Research and Development Program of China under Grant 2023YFB3107402; and in part by Hangzhou Leading Innovation and Entrepreneurship Team under Grant TD2020003. *(Corresponding author: Li Lu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) in Zhejiang University.

Li Lu, Meng Chen, Zhongjie Ba, Feng Lin, Jinsong Han, and Kui Ren are with the State Key Laboratory of Blockchain and Data Security, School of Cyber Science and Technology, and the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310007, China (e-mail: li.lu@zju.edu.cn; meng.chen@zju.edu.cn; zhongjieba@zju.edu.cn; flin@zju.edu.cn; hanjinsong@zju.edu.cn; kuiren@zju.edu.cn).

Jiadi Yu and Yanmin Zhu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jiadiyu@sjtu.edu.cn; yzhu@sjtu.edu.cn).

Digital Object Identifier 10.1109/TNET.2024.3403839

communication range, leading to the activity semantics leakage. Combining with the fact that many household WiFi-connected devices suffer from imperceptible compromising attacks [22], such a leakage introduces more significant threats when we enjoy the convenience brought by WiFi sensing techniques. Representative threats include leaking your private information (such as your gender or behavioral identity) to malicious neighbors, causing further crimes against your property and even life.

Along this direction, this work aims to further investigate the feasibility of eavesdropping on the omni-directional broadcasting signal to retrieve individual identity for impersonation attacks by pervasive WiFi infrastructures, on the basis of our previous work. In this attack, we follow the same attack scenario, i.e., an adversary has compromised one of widely deployed WiFi infrastructures, which could be placed in any location, as demonstrated to be common recently [22]. Using the compromised device, the adversary eavesdrops on the leaked signals interfered with by victim user. The adversary can further employ the signature underlying the leaked signals to impersonate the legitimate user to conduct sensitive operations, such as online shopping, which is similar to compromising the voiceprint recognition for shopping on Amazon Echo. Different from our previous work facing the three challenges, including *single device*, *unknown device location* and *black-box attack*, to realize the impersonation attack, we need to further validate the feasibility of following problems. *Finer-grained features in converted signals*: compared with activity semantics retrieval, the impersonation attack relies on finer-grained individual behavioral uniqueness, which is more easily interfered with by ever-existed noises, indicating its difficulty with converted signals. *Signal injection for impersonation*: to impersonate a legitimate user, the adversary should be able to access the target device or API to inject the signals, which is unnecessary for activity surveillance.

In this paper, we extend the attack capability of *ActListener* from activity surveillance to impersonation attacks, which could eavesdrop on the airborne WiFi signal carrying legitimate user's activity semantics and individual identities for the attacks. Specifically, in this attack, an adversary either has compromised one of the WiFi infrastructures in the user's space in advance, or is able to place his/her own receiver in the user space. When the victim user performs an activity between the transmitter and receiver following the similar process of WiFi sensing systems, his/her transmitter continuously transmits the omni-directional broadcasting WiFi signals, which interact with the user's body, and are received by the legitimate receiver as well as the compromised device. From eavesdropped signals, *ActListener* follows the previous design [21] to detect signal segments induced by valid user activity, and estimates the relative locations of the victim user and his/her receiver from the adversary's device for further signal conversion. Then, *ActListener* converts eavesdropped signals to that received by the legitimate receiver based on their propagating signal models to obtain the activity semantics uniqueness underlying WiFi signals. To extract noise-resilient individual behavioral uniqueness from converted CSI of WiFi signals, we further add user identification models into the substitute model set for training the signal pattern calibration generative model. Finally, we implement several signal injection methods to transfer the converted signals to WiFi-based user identification for impersonation attacks. Extensive experiments demonstrate that *ActListener* is robust and efficient to recover the signal carrying the legitimate user's behaviors and individual uniqueness, for enabling the activity surveillance and impersonation with any compromised WiFi infrastructure.

We highlight our contributions as follows:

- We demonstrate *ActListener*, which could eavesdrop on WiFi signals induced by user activities for retrieving not only coarse-grained activity semantics, but also fine-grained individual identity without prior knowledge of the recognition model and device locations.
- We design an activity modeling-based signal conversion method, which could recover received signals from a legitimate receiver only based on eavesdropped signals without acquiring model details in advance.
- We extend the generative model-based signal calibration approach, which could resist the always-existed noises in CSI of over-the-air WiFi signals to recover a robust signal for both activity surveillance and impersonation attacks.
- We conduct experiments in real environments and the results show that our recovered signal achieves an 88.4% average α-similarity with originally received signals and over 80% accuracy in activity semantics retrieval and impersonation.

II. Preliminary

In this section, we introduce the system model of WiFi sensing, then illustrate the threat model of WiFi-based activity surveillance, and show its feasibility study.

A. System Model

WiFi sensing has been widely investigated for natural Human-Computer Interactions (HCIs) these years [4], [5], [6], [7], [8], [9], [10], [23], [24]. Even with the increasing requirement for security and privacy in IoT environments, WiFi-based user identification has also been widely investigated [12], [13], [17], [18], [25], [26], including the access control of interactive systems, and house-wide security protection. The basic idea of WiFi sensing is to capture the human motions using WiFi signal and explore the corresponding signal patterns to recognize the human activities (e.g., coarse-grained daily activities, fine-grained interactive finger gestures). Taking advantage of the finer-grained sensing capability of Channel State Information (CSI) [27], WiFi signals could even capture the subtle behavioral uniqueness of distinct individuals and contribute to realizing device-free user identification.

Principle. WiFi sensing usually consists of data collection, signal processing, feature extraction, classification model training, and functional classification (i.e., activity recognition or user identification and spoofer detection). Data collection extracts CSI of WiFi signals interfered by human body with commercial network interface card for behavior representation, i.e.,

$$H(f) = Y(f)X(f)^{-1}, \tag{1}$$

where $Y(f)$ is the received signal, $X(f)^{-1}$ is the inverse of transmitted signal $X(f)$. In human activities, the human body interferes with WiFi signals, leading to the change of channel properties, which are exhibited in CSI amplitude and phase. Then, signal processing eliminates unrelated interference and unnecessary redundancy from collected raw data. Representative processes include noise elimination, subcarrier selection, signal normalization, etc. Furthermore, feature extraction and classification models training module employs various machine learning models (e.g., sparse approximation [26], decision tree [25], autoencoder [17], recurrent neural network [12], [13], support vector machine [12], [13], [18]). Finally, the functional classification integrates all the three previous modules to realize specific functions for the sensing system, such as retrieving the activity semantics for HCIs with IoT devices for activity recognition, or achieving privacy protection for IoT devices for user identification and spoofer detection. Compared with vision-based activity recognition or user identification, the number of WiFi sensing channels is much lower. Hence, various relative locations and orientations induce significant interference in accurate activity recognition or user identification.

Security Analysis. Actually, such a WiFi sensing system is inborn with the vulnerability to various attacks, leading to probable compromise of the system. (1) Due to the broadcasting manner, the WiFi signals carrying the activity semantics or individual behavioral uniqueness could be captured by an arbitrary receiver in the sensing area, which leads to the probable leakage of user information. (2) Directly eliminating the interference and redundancy without cross-checking probably reduces the difficulty of attacking the system. For example, the signal normalization process normalizes CSI to the range of $[0, 1]$, which releases the requirements of estimating the exact amplitude value for the attack. (3) Though the variety of machine learning models may contribute to resisting some attacks (e.g., adversarial attack), the black-box attacks (i.e., a successful attack without the knowledge of model details) bypassing the models could introduce more imperceptible and severe threats in practice.

B. Threat Model

With the convenience brought by WiFi sensing, the privacy compromising threat is also introduced. Once an arbitrary receiver in the sensing area is compromised, due to the broadcasting manner, WiFi signals carrying the activity patterns could be captured by the receiver, which leads to the probable leakage of daily activities. However, since WiFi is used to realize communication for most users intuitively, such a sensing or even surveillance capability of WiFi does not raise user awareness, leading to even more severity than recently revealed surveillance camera leakage [28], [29]. Such a privacy compromising threat also occurs when the adversary is able to place his/her own receiver in the user space. Such a scenario is common, e.g., the user space is an open office, or the adversary is actually a curious friend or relative of the victim user.

We assume a victim user employs the WiFi sensing (as mentioned in Section II-A) providing the natural HCIs with

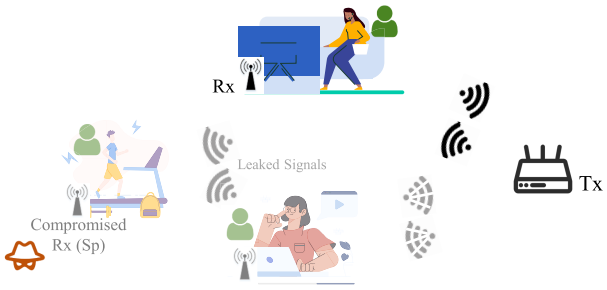


Fig. 1. Illustration of threat model.

smart appliances (e.g., smart TV) in an indoor environment (e.g., his/her home or office) for activity recognition or user authentication. Although the user may also adopt different recognition and authentication systems based on other modalities (e.g., image, millimeter wave, and IMU), we focus on WiFi sensing since it provides a natural and continuous interaction experience, which is more user-friendly. In this case, a WiFi gateway serves as the transmitter (Tx) and other appliances integrate the receivers (Rx) respectively. Considering the strong diffraction property of WiFi signals, the most reliable sensing area concentrates on the first Fresnel Zone [30], i.e., the user usually performs the interactive gestures on the LoS between transmitter and receiver antennas for a robust response. Hence, in this work, we follow a similar system setup, i.e., the victim user performs the gestures in the middle between transmitter and receiver antennas, and other persons act outside the first Fresnel Zone. We assume the receiver extracts CSI from WiFi signals and runs machine learning algorithms to perform activity recognition or user identification. Moreover, since almost all existing works [4], [5], [6], [7], [8], [9], [12], [13], [17], [18], [25], [26] utilize CSI amplitude to explore activity semantics or user behavioral uniqueness for further recognition or identification, the targeted signal patterns during eavesdropping for the attack are based on CSI amplitude of WiFi signals.

Fig. 1 shows the threat model. In the attack, we assume one of the victim's Rx is compromised, which is not rare in real world [22] due to the lack of users' security awareness. Such a compromised WiFi receiver could be turned into an eavesdropping device (Sp) to leak the user privacy, as the traditional surveillance camera acts. By sniffing WiFi signals with the compromised device, the adversary eavesdrops on the activity semantics and behavioral uniqueness of the victim user, which serves as the prerequisite for privacy retrieval or impersonation. To avoid raising the victim user's awareness, the adversary could only compromise the receiver in the digital domain, i.e., without any physical access to the device. In this case, the adversary can either be within the range of targeted WiFi or remotely communicate with the compromised device via Internet. Hence, the adversary has no prior knowledge of the relative positions between Tx and Rxes. Meanwhile, the relative positions between Sp and Rxes are also unknown to the adversary, because the receiver may be compromised remotely. In these cases, both Tx-Rx and Tx-Sp pairs are not always within LoS for sensing. The only way to turn the compromised receiver into an eavesdropping device is to infer the human activities or user identity from

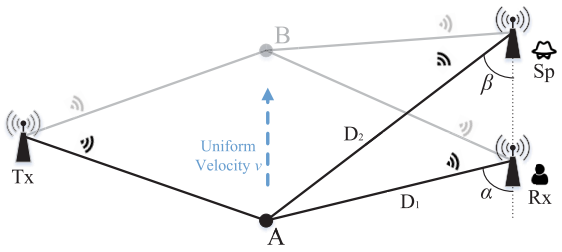


Fig. 2. Illustration of a mass point affecting signals of different WiFi receivers.

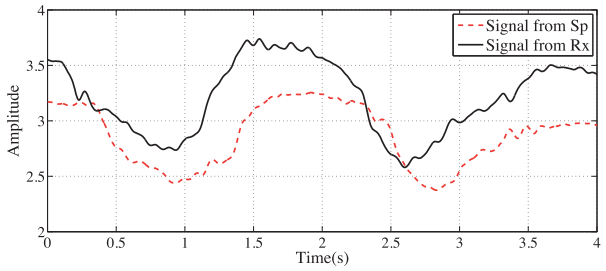


Fig. 3. CSI amplitudes of received signals from Rx and Sp.

the captured omni-directional WiFi signals that interacted with users. Considering the smart device may not always maintain a continuous transmission for sensing, we assume the continuous connection between Tx and Rxes should sustain to sense a complete user activity for human-computer interaction. For example, the duration can be 0.5s for slide gesture while 1s for zigzag gesture. Hence, we select the longest duration (around 2s) of a complete interacted gesture as the required duration of continuous transmission. Also, the adversary has no prior knowledge of activity recognition model or user identification details, including the structure and detailed parameters, i.e., a black-box attack. As a result, the only way for activity surveillance or impersonation is to recover the original signals received by the legitimate Rx and fed them into various recognition or identification models.

C. Feasibility Study

To realize the activity surveillance with any receiver, we first analyze the propagating WiFi signals theoretically and conduct an experimental study to validate its feasibility.

To model the human activity based on WiFi signals, we start by simplifying the human body as a mass point to derive its effect of linear movement on WiFi signals. The mass point mainly simulates the significant movement of human trunk, while ignoring other subtle movements of limbs, to model its impact theoretically. As illustrated in Fig. (2), the mass point moves in a linear way (i.e., from 'A' to 'B'), and a WiFi system containing a Tx and an Rx is deployed for activity recognition. The signal is first transmitted from Tx, then interacts with the targeted mass point, and is finally received by Rx. At time t, the received signal $Y(f,t)$ is formulated as

$$Y(f,t) = a(f,t)e^{-j2\pi\frac{D(t)}{\lambda}}, \tag{2}$$

where f is the subcarrier frequency of WiFi signals, $a(f,t)$ is the propagating attenuation coefficient and $D(t)$ is the distance between mass point and Rx. Note that received signals consist of scattering signals from the targeted mass point, so the mass point could be regarded as a relay signal source. Toward this end, $D(t)$ in Eq. (2) is the distance between mass point and Rx, instead of the propagating distance of signals. Furthermore, according to the inverse square law [31], we could formulate the attenuation coefficient with signal propagation distance:

$$a(f,t) = \frac{k \cdot a(f,0)}{D(t)^2}, \tag{3}$$

where k is the proportionality coefficient. Also, the transmitted signal $X(f)$ could be regarded as the received signal at time 0, i.e., $X(f) = Y(f,0) = a(f,0)e^{-j2\pi f\cdot\frac{0}{\lambda}} = a(f,0)$. Hence, with Eq. (2) and (3), we have

$$Y(f,t) = a(f,0)\frac{k}{D(t)^2}e^{-j2\pi\frac{D(t)}{\lambda}} = X(f)\frac{k}{D(t)^2}e^{-j2\pi\frac{D(t)}{\lambda}}. \tag{4}$$

Combined with Eq. (1), the CSI is derived as

$$H(f,t) = \frac{k}{D(t)^2}e^{-j2\pi\frac{D(t)}{\lambda}}. \tag{5}$$

Eq. (5) indicates that the difference in the target-receiver distance is exhibited in CSI $H(f,t)$. This principle motivates us to utilize the location geometry between Rx and a spoofing receiver (Sp) to realize the signal conversion of CSI from different receivers for eavesdropping.

To validate the feasibility of activity surveillance with any receiver in the target's sensing space, we conduct an experiment simulating the scenario in Fig. 2. In the experiment, a volunteer is recruited to simulate the mass point. The volunteer moves along the perpendicular bisector of the Tx-Rx connection without complex behaviors. On the other hand, an Sp is placed directly facing Rx to eavesdrop on CSI induced by the moving volunteer. Fig. 3 shows CSI amplitudes of received signals from Rx and Sp. We can observe that though Rx and Sp are placed in different locations, their received CSI amplitudes exhibit similar trends, because they are induced by the same behavior of the volunteer (i.e., moving in a linear trajectory). This result reveals the underlying relationship between signals received by Rx and Sp. Specifically, the volunteer moves from an initial position to 'A' as shown in Fig. 2. The distance and angle of 'A' from Rx are D_1 and α respectively, and that from Sp are D_2 and β. In this case, the CSI amplitudes of Rx and Sp become

$$\|H_i(f,t)\| = \|\frac{k_i}{(D_j)^2}e^{-j2\pi\frac{D_j}{\lambda}}\| = \frac{k_i}{(D_j)^2}, \tag{6}$$

where $i \in \{Rx, Sp\}$, $j \in \{1,2\}$ respectively. With the location geometry between Rx and Sp, we have $D_2 \sin\beta = D_1 \sin\alpha$. Substituting this equation into $\|H_{Rx}(f,t)\|$, we obtain

$$\begin{aligned}\|H_{Rx}(f,t)\| &= (\frac{\sin\alpha}{\sin\beta})^2\frac{k_{Rx}}{(D_2)^2} \\ &= (\frac{\sin\alpha}{\sin\beta})^2 \cdot \frac{k_{Rx}}{k_{Sp}}\|H_{Sp}(f,t)\|.\end{aligned} \tag{7}$$

Eq. (7) validates the underlying relationship between CSIs from Rx and Sp, consistent with experimental results.

The result and analysis validate the feasibility of converting CSI amplitudes of airborne WiFi signals received by diversely

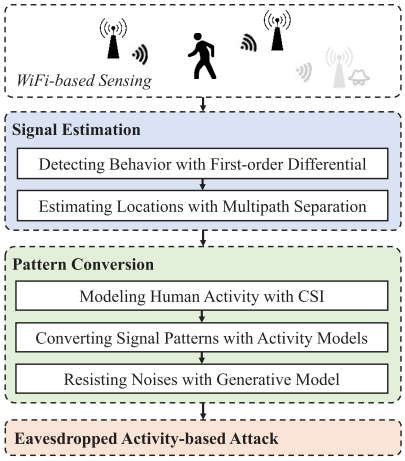


Fig. 4. Architecture of *ActListener*.

located receivers. This encourages us to realize the signal conversion for recovering the signal received by the legitimate user's device from sniffed signals to eavesdrop on the behavioral uniqueness underlying the signals.

III. Design of *ActListener*

In this section, we present design details of the imperceptible activity surveillance attack *ActListener*.

A. Overview

To reveal the threat of activity surveillance by any compromised WiFi infrastructure in the victim's space, we demonstrate *ActListener*, which requires no direct physical access to the victim user's devices and the prior knowledge of the device location and activity recognition models. Fig. 4 shows the architecture of *ActListener*. In the attack, the adversary has compromised one of the WiFi infrastructures in the victim user's space, which widely exists recently as mentioned in Section II-B. When the victim user acts in WiFi coverage, the Tx continuously transmits the omni-directional broadcasting WiFi signals, which interact with the user's body, and are received by the legitimate Rx and the adversary's Sp. From eavesdropped signals by Sp, *ActListener* aims to recover the signals received by the legitimate Rx for further activity recognition. *ActListener* first detects signal segments induced by a valid user activity by first-order differentials, and then estimates relative locations of the victim user and his/her devices from Sp to provide the parameters for further signal conversion. After that, *ActListener* models received signals of Rx and Sp respectively based on CSI amplitudes, and further converts Sp's received signals to that of Rx based on the models. Finally, a generative model is employed to calibrate the converted signals to resist the noises in CSI of WiFi signals. By the signal conversion, *ActListener* could unify the WiFi signals induced by the victim user into the same coordinate system, and further fed them into the corresponding activity recognition model to realize the activity surveillance.

B. Signal Estimation

To realize the signal conversion in Section II-C, *ActListener* first needs to capture the signal interacting with the legitimate user's behaviors, and estimate the relative locations between the adversary's and victim user's devices.

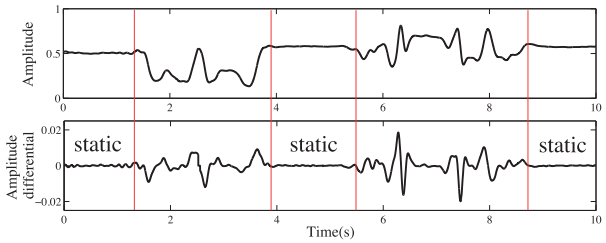


Fig. 5. CSI amplitudes and their first-order differentials.

1) Detecting Activity With First-Order Differential: When *ActListener* activates the activity surveillance, the system continuously receives the omni-directional broadcasting WiFi signals. Note that Sp has already connected with Tx in the attack, so other wireless signals around could not interfere with the activity surveillance. The signals that interacted with both user behavior and static environments are in the received signals. Hence, *ActListener* should first detect the signal segments that correspond to each valid activity for further conversion.

Fig. 5 shows the received CSI amplitudes of WiFi signals interacted with a series of activities. We can observe that there exists a sudden variance in CSI amplitudes at the start and end of an activity. This is because human activity changes the position of human body, affecting the propagation path of WiFi signals. Hence, the CSI amplitude would change accordingly, which is consistent with Eq. (5). This observation inspires us to employ the threshold-based method to detect the activity in received CSI amplitudes. However, due to the probable difference in the human body positions before and after an activity, the CSI amplitude before and after the activity would be different, leading to difficulty in determining the threshold for activity detection.

Taking a deep look at the top part of Fig. 5, we can also find that the CSI amplitudes remain relatively stable when no user acts in the environment, while that under human activities exhibit significant variances. To exploit this property, we derive the first-order differential of CSI amplitudes, as shown in the bottom part of Fig. 5. It can be observed from the figure that the amplitude differentials under static environment are limited to a narrow range around the value of zero, while that under human activity shows significant fluctuation. Hence, we employ a sliding window to detect whether the values of all signal points are within a threshold, so as to detect the start and end of the signal segment corresponding to independent human activity. The sizes of sliding window and threshold could be determined by empirical study, which are set as $500ms$ and 0.4×10^{-3} in our attack, respectively.

2) Estimating Locations With Multipath Separation: Revisiting the attack principle in Section II-C, the signal conversion relies on the relative location between Rx and Sp, indicating the requirement of location estimation for *ActListener*. The relative locations require to be estimated include their distances and angles, which could be measured by Time of Flight (ToF) and Angle of Arrival (AoA) joint estimation. However, the joint estimation introduces significant cumulative

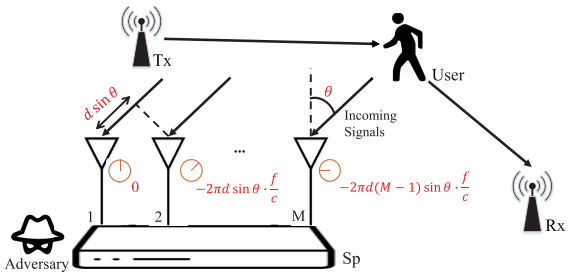


Fig. 6. Illustration of AoA estimation with CSI phase shifts.

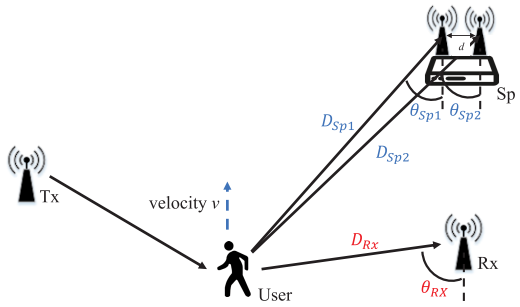


Fig. 7. Illustration of location estimation with multiple estimated AoAs.

errors, significantly degrading the signal conversion performance. To accurately estimate relative locations between Sp and a target (including Tx, Rx, and moving body), we turn to separate the multipath signals propagating from different locations via the widely-used multipath profiling algorithm, i.e., the high-resolution MUlti SIgnal Classification (MUSIC) [32], [33], [34], on commercial WiFi infrastructures.

Suppose the adversary carries Sp with M receiver antennas arranged as a linear array with an interval of d, as shown in Fig. 6. A signal is transmitted from Tx, then propagates through multiple paths by interacting with ambient objects (such as moving body, wall), and is finally received by the Sp. Hence, the angle of directly propagating signal from the target to Sp, i.e., the AoA of Sp's incoming signal θ, could be utilized to estimate the relative location between the target and Sp. We estimate the AoA θ from CSI phase of WiFi signals received by Sp. Specifically, *ActListener* sets 1^{st} antenna's CSI phase as the basis, and derives the relative phases of other antennas accordingly. Hence, except the 1^{st} antenna's CSI phase being 0, the relative phase of m^{th} antenna is $-2\pi d(m-1)sin(\theta)\cdot\frac{f}{c}$, where f and c are the signal frequency and speed respectively. Hence, the Inphase-Quadrature (IQ) signal of CSI phase received by the m^{th} antenna is

$$\phi^m(\theta) = e^{-j2\pi d(m-1)\sin(\theta)\cdot\frac{f}{c}}. \tag{8}$$

Based on the IQ signal, we construct the steering matrix as $\mathbf{A} = [\vec{a}(\theta_1) \quad \vec{a}(\theta_2) \quad \cdots \quad \vec{a}(\theta_L)]$, where $\vec{a}(\theta_i) = [1 \quad \phi^1(\theta_i) \quad \cdots \quad \phi^{M-1}(\theta_i)]^T$ is the steering vector of i^{th} path. Hence, the signal $\mathbf{X}$ could be modeled as

$$\mathbf{X} = [\vec{x_1} \quad \vec{x_2} \cdots \vec{x_N}] = \mathbf{AF} + \mathbf{N}, \tag{9}$$

where N is the number of subcarriers, $\vec{x_1}, \cdots, \vec{x_N}$ are the received signal vectors each subcarrier, $\mathbf{F} = [\vec{F_1}, \cdots, \vec{F_N}]$ is the attenuation matrix including the attenuation coefficients of each subcarrier in each path, and $\mathbf{N}$ is the noise matrix.

Based on the signal model, *ActListener* further constructs the pseudo-spectrum for AoA estimation. During the eavesdropping, the adversary's Sp could receive the CSIs organized as a matrix, in which each row represents the signals received by different antennas, and each column indicates that by different subcarriers. In the matrix, each column corresponds to the received signal vector in the signal model of Eq. (9). According to the correspondence, *ActListener* could estimate the AoAs as long as the steering matrix $\mathbf{A}$ could be derived from the CSI matrix. Specifically, we first perform eigenvalue decomposition of the covariance matrix $\mathbf{R} = \mathbf{XX}^H$, where $\mathbf{X}^H$ is the conjugate transpose of $\mathbf{X}$. Then, the D eigenvectors with the largest D eigenvalues are selected to construct the signal subspace. The rest M - D eigenvectors form the noise subspace $\mathbf{E}_N$. With orthogonality between the noise subspace and signal subspace [35], *ActListener* derives the pseudo-spectrum as

$$P(\theta) = \frac{1}{\vec{a}(\theta)^T\mathbf{E}_N\mathbf{E}_N^H\vec{a}(\theta)}. \tag{10}$$

Eq. (10) exhibits the pseudo signal's intensity under different AoAs. Hence, the angle θ with a local maximum value in the pseudo-spectrum $P(\theta)$ is regarded as the AoA from the target, i.e., $\theta = \arg\max_{\theta}^{local} P(\theta)$.

Except for AoAs, *ActListener* needs to further estimate the distance between Sp and the target so as to estimate their relative location. We estimate the distance by estimating the ToAs with a modified MUSIC algorithm [36]. Different from traditional MUSIC, we formulate the CSI of each subcarrier as a function of ToA, i.e., $h(f) = \sum_{i=1}^{N} a(f)e^{-j2\pi\Delta t} + j\phi(f)$, where Δt is the ToA, $\phi(f)$ is the phase shift of the subcarrier. Based on it, we could re-formulate the mode vectors, signal matrix $\mathbf{X}$ and the attenuation matrix $\mathbf{F}$. Considering search space difference between AoAs (i.e., $-\pi \sim \pi$) and ToAs (i.e., $-\infty \sim \infty$), we further set the search space as $[-1/f_\delta, 1/f_\delta]$. To further improve the resolution, the received signals from multiple receiver antennas are involved, as illustrated in Fig. 7. Since the CSIs from different antennas are linearly independent while their ToA difference is subtle (i.e., $D_{Sp1} \approx D_{Sp2}$ in Fig. 7), we could combine CSIs from closely placed antennas to enhance the resolution of ToA estimation.

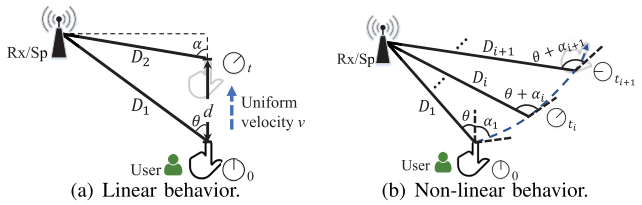
C. Pattern Conversion

After the location between Sp and other targets are estimated, *ActListener* could realize the pattern conversion for the activity surveillance.

1) Modeling Human Activity With CSI: Although Section II-C has demonstrated the feasibility of employing any compromised WiFi infrastructure for activity surveillance by pattern conversion, the impact of human activity on WiFi signals in real environments is different from that of an ideal mass point due to multipath effects. Theoretically, considering the connection between Sp/Rx and Tx during activity recognition, the CSI consists of two parts only, i.e.,

$$H(f,t) = \frac{k}{D(t)^2}e^{-j2\pi\frac{D(t)}{\lambda}} + N, \tag{11}$$

where the first term is Eq. (5) indicating the component induced by human activity, and N is a constant representing



(a) Linear behavior. (b) Non-linear behavior.

Fig. 8. Modeling a human activity with CSI of WiFi signals.

the component caused by static part and noises. Consider a user performs a linear gesture as shown in Fig. 8(a), in which the distance between a user and receiver (i.e., the legitimate Rx or illegal Sp) is D_1 initially. His/her body part then moves with a sufficiently short distance d at time t (the moving velocity is considered as a constant v), in which the distance between the user and receiver is D_2. With the geometry, we have

$$D_2 \sin\alpha = D_1 \sin\theta, \qquad D_2 \cos\alpha = D_1 \cos\theta - d. \tag{12}$$

Combining the squares of the two equations above, we can derive:

$$\begin{aligned}(D_2)^2 &= (D_1)^2 \sin^2\theta + (D_1\cos\theta - d)^2 \\ &= (D_1)^2(1 + (\frac{d}{D_1})^2 - 2\frac{d}{D_1}\cos\theta).\end{aligned} \tag{13}$$

In real situations, when the user performs a linear gesture, the whole body (e.g., arm), instead of only a body part (e.g., hand), moves, so such movements could not be regarded as a mass point moving. To model the linear human behavior on CSI in practice, we further divide the whole body into infinite mass points, and then the overall CSI of the whole body is the integral of all individual CSIs for each mass point in the moving body, i.e.,

$$\begin{aligned}H_T(f,t) &= \int_0^t H(f,t) v d\Delta t + N \\ &= \int_0^t \frac{kv \cdot e^{-j2\pi\frac{D_2}{\lambda}}}{(D_1)^2(1 + (\frac{v\Delta t}{D_1})^2 - 2\frac{v\Delta t}{D_1}\cos\theta)} d\Delta t + N.\end{aligned} \tag{14}$$

Due to the unseen value of N and integral operation, this model could be hardly employed for pattern conversion. Hence, we further derive the differential of Eq. (14), i.e.,

$$dH_T(f,t) = \frac{kv}{(D_1)^2(1 + (\frac{vt}{D_1})^2 - 2\frac{vt}{D_1}\cos\theta)}, \tag{15}$$

where $dH_T(f,t) = \frac{dH_T(f,t)}{d\Delta t}$ for simplicity. Using Eq. (15), *ActListener* could model the linear human behavior with CSI of WiFi signals.

Except for linear human behavior, both daily activities and interactive gestures contain non-linear behaviors. We further derive the model of non-linear behaviors on CSIs. Usually, a non-linear behavior could be regarded as a combination of multiple linear behaviors. Hence, we divide a non-linear behavior into multiple segments, each of which is short enough to be approximately regarded as a linear behavior. Fig. 8 illustrates the modeling of a non-linear behavior by segmenting it into multiple linear behaviors. Based on Eq. (12), for each segment, we have

$$D_{i+1}^2 = D_i^2(1 + (\frac{d}{D_i})^2 - 2\frac{d}{D_i}\cos(\theta + \alpha_i)), \tag{16}$$

where d is the moving distance of the behavior, D_i and D_{i+1} are the distances from the receiver of i^{th} and $(i+1)^{th}$ behavior segments respectively. Similarly, the differential model $dH(f, t_{i+1})$ is

$$dH(f,t_{i+1}) = \frac{kv}{D_i^2(1 + (\frac{d}{D_i})^2 - 2\frac{d}{D_i}\cos(\theta + \alpha_i))}, \tag{17}$$

where t_{i+1} is the time index of the $(i+1)^{th}$ behavior segment. Note that the formulations of Eq. (15) and (17) are actually consistent. Therefore, *ActListener* could model both linear and non-linear behaviors in the same manner.

2) Converting Signal Patterns With Activity Models: Based on the activity modeling, *ActListener* could formulate the targeted victim user's activity from Rx's and Sp's perspectives respectively, and realize the pattern conversion between them.

Suppose a user performs an activity in the WiFi coverage, as shown in Fig. 7. Based on Eq. (15), *ActListener* models the activity from the two perspectives:

$$dH_i(f,t) = \frac{k_i v}{(D_i)^2(1 + (\frac{vt}{D_i})^2 - 2\frac{vt}{D_i}\cos\theta_i)}, \tag{18}$$

where $i \in \{Sp, Rx\}$, $dH_{Sp}(f,t)$ and $dH_{Rx}(f,t)$ are the activity models of CSI received by Sp and Rx respectively, t is the time index, v is the moving velocity of the activity, θ_{Sp} and θ_{Rx} are the angles between the user's moving body and the receivers (i.e., Sp and Rx) respectively. During the attack, the adversary could directly observe the value of Sp's CSI amplitudes (i.e. $H_{Sp}(f,t)$) by eavesdropping on the broadcasting WiFi signals, so the value of $dH_{Sp}(f,t)$ is known to *ActListener*.

The task of pattern conversion is to recover the expression of $dH_{Rx}(f,t)$ based on $H_{Sp}(f,t)$ so as to unify the legitimate Rx's received signal $H_{Rx}(f,t)$ into the same coordinate system with that from the legitimate receiver for activity surveillance. Specifically, we perform polynomial expansion on Sp's signal model $dH_{Sp}(f,t)$ and obtain the approximation $dH_{Sp}(f,t) \approx \frac{k_{Sp}v}{D_{Sp}^2}(1 + \frac{2v}{D_{Sp}}\cos\theta_{Sp} \cdot t)$. And we could obtain the constant and first-order coefficient

$$a_1 = \frac{k_{Sp}v}{D_{Sp}^2}, \quad a_2 = \frac{k_{Sp}v}{D_{Sp}^2} \cdot \frac{2v}{D_{Sp}}\cos\theta_{Sp}. \tag{19}$$

In Eq. (19), both D_{Sp} and θ_{Sp} could be obtained with the location estimation method in Section III-B.2. On the other hand, the two coefficients a_1 and a_2 could be derived from the Sp's received CSI $H_{Sp}(f,t)$. Hence, Eq. (19) could be treated as an optimization problem with two unknown variables. By solving the problem, *ActListener* could obtain the moving velocity of the human behavior v and the Sp coefficient k_{Sp}.

Revisiting Eq. (18), we find there remain three parameters, i.e., k_{Rx}, D_{Rx} and θ_{Rx}, before enabling the pattern conversion between Sp and Rx. k_{Rx} could be obtained by enumerating all relative locations between Sp and Rx with empirical studies. On the other hand, D_{Rx} and θ_{Rx} are related to the relative

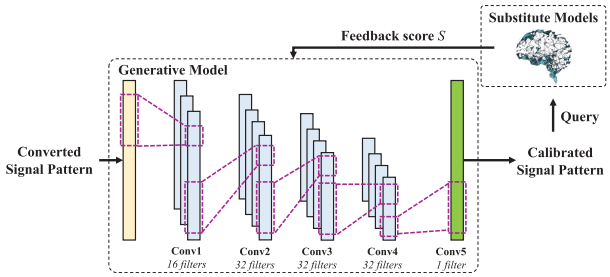


Fig. 9. Illustration of generative model for signal calibration.

position of Rx from Sp. Since *ActListener* has already estimated the location of Rx and the victim user as mentioned in Section III-B.2, the values of D_{Rx} and θ_{Rx} could be obtained with their geometry.

With obtained parameters, *ActListener* realizes the pattern conversion between Sp and Rx. Specifically, *ActListener* first derives the signal model received by Sp $dH_{Sp}(f,t)$ by the differential operation, i.e., $dH_{Sp}(f,t) = H_{Sp}(f,t+1) - H_{Sp}(f,t)$, and obtains k_{Sp} and v by solving Eq. (19). Then, combined with v, estimated D_{Rx} and θ_{Rx}, and k_{Rx}, *ActListener* further derives the signal model $dH_{Rx}(f,t)$ with Eq. (18). To recover the received signal of legitimate Rx, *ActListener* further performs the integral operation on $dH_{Rx}(f,t)$, i.e., $H_{Rx}(f,t+1) = dH_{Rx}(f,t) + H_{Rx}(f,t)$.

3) Resisting Noises With Generative Model: Theoretically, *ActListener* is able to recover the signal received by the legitimate Rx after the activity modeling-based signal conversion. However, since the adversary probably compromises a single device only, the angular resolution is insufficiently high. Moreover, considering the ever-existing noises in CSI of WiFi channels in practice, the converted signal is not robust enough for practical activity surveillance, and even worse for impersonation attack. To resist the noises in CSI, we propose a generative model-based signal calibration to recover noise-resistant signals.

The generative model takes the converted signal as input, and outputs the noise-resistant signals accordingly. Specifically, the generative model G is basically a 5-layer Time-Delay Neural Network (TDNN) [37], as illustrated in Fig. 9. The network is designed with five 1-D convolution (Conv) blocks [38]: the first block is to unfold the input converted signal to feature-representation with 16 feature filters (channels), and the last block is to wrap the features into 1 channel to produce the output noise-resistant signals. Between them, three identical Conv blocks are designed to learn the deep representation of features. Within each block, a convolutional layer with 1×3 convolution kernels and 32 feature filters are implemented, followed by a Batch-Normalization (BN) [39] layer, and then a Leaky ReLU (LReLU) [40] layer as the activation function.

Except for the network architecture, we further design a loss function guiding the training optimization to generate noise-resistant signal representations. As mentioned in Section II-B, the adversary has no prior knowledge about the target activity recognition system on the receiver device. To involve the noise impact into calibrated signals without access to the target activity recognition and user identification models, we introduce several local activity recognition and user identification models as the substitutes, including CNN-based and LSTM-based models, which have been used in previous WiFi-based activity recognition and user identification works. As shown in Fig. 9, given the generated signals, the substitute models provides the query score $S(S \in (0,1))$ as a valuable feedback to optimize the generated signals. Considering the intrinsic transferability of the activity recognition and user identification models, the higher the score S is, the more accurate and robust the generated signal is. Since a legitimate signal for bypassing the system always involves the noises in CSI, the generative model could employ the query scores S for training to formulate the calibrated signal with the noise impact in practical scenarios. Following the principle, the loss function for model training is formulated as $\mathcal{L} = -y \log S$, where y is the ground truth (user activity or identity label). Considering the finer-grained features needed for user identification, the training is actually organized by a two-step manner, i.e., the model is first trained with the activity recognition model's feedback and then with the user identification model's score.

We build two datasets to train the generative model. The first dataset collects the real-world CSI data, which is collected by the Sp when a legitimate user acts for activity recognition. In practice, we only collect 20 samples in this dataset, including two simple behaviors, i.e., push and pull, each for 10 samples. The second dataset contains only augmented samples constructed from raw samples in the first dataset. For each sample in the first dataset, we add different levels of Guass noises into the sample to create 50 new augmented samples. In total, the second dataset contains 1,000 data samples for training. Based on the datasets, the generative model could be trained. In particular, the raw samples are first sent to the system to calculate the converted pattern as described in Section III-C.2. Then, each converted pattern related to a behavior is re-vectorized as a vector with a length of $3N$, where N is the sampling rate of CSI. The rationale is that the processes of most user behaviors last for less than $3s$. If a user activity lasts for $\hat{t}$ seconds ($\hat{t} < 3$), we could add $\lfloor (3-\hat{t})N \rfloor$ zero samples at the end of the signal to align its length to $3N$, so that the generative model could treat all the training samples in a parallel form. During the training phase, we adopt an Adam optimizer whose learning rate follows the cosine annealing schedule from 10^{-3} to 10^{-5}, and a weight decay factor of 0.9 and early stopping is used for alleviating overfitting. These parameters are selected according to empirical study. We iterate the training procedure to update the model parameters continuously according to the feedback score until the loss function converges.

During the optimization, the generative model continuously queries the substitute system and learns to resist noise interference for better recognition and identification performance. Under the guidance of the substitute system, most insignificant information is discarded and valuable patterns are remained to generate more robust signals. After optimization on thousands of converted signals with different noises, the generative model is endowed with satisfactory denoising ability. Based on such a well-trained generative model, *ActListener* could reconstruct noise-resistant signals from the raw converted signals.

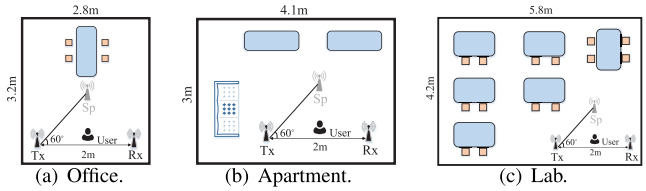
D. Eavesdropped Activity-Based Attack

After the pattern conversion, *ActListener* could obtain the behavioral patterns underlying WiFi signals for the adversary to launch corresponding activity surveillance and impersonation attacks.

1) Activity Semantics Extraction: To realize the activity surveillance, the adversary further needs to extract the activity semantics from the signals. Since the adversary could not physically access the legitimate user's device, *ActListener* turns to query the cloud-based models with the converted signal patterns for activity semantics extraction. Multiple prevalent activity recognition models that can be accessed publicly with the converted signal patterns. We employ two different strategies in terms of the permission that the adversary could obtain. Specifically, if the adversary is a curious one, e.g., the victim user's friend or relative, he/she may be able to connect to the user's LAN. In such a case, the adversary sniffs the packets sent from the legitimate user's terminal, and retrieves the destination IP address that correlates to the cloud-based model. After that, *ActListener* reconstructs the packet containing the generated signal pattern as the payload and the destination IP address, and queries the targeted cloud-based model for the activity semantics extraction. On the other hand, if the adversary is a malicious one, i.e., he/she could not obtain the sensitive packet inner the LAN, the adversary turns to query multiple prevalent activity recognition models that could be accessed publicly as the substitute.

2) Signal Injection for Impersonation: Except for extracting activity semantics for surveillance, the adversary could further inject the converted signal patterns into legitimate devices for impersonation. *ActListener* eavesdrops on the broadcasting WiFi signals interacting with the victim user's behavior, and then converts the signal into that received by the legitimate Rx. Hence, the converted WiFi signals intrinsically embed the individual behavioral uniqueness. Such an eavesdropped and converted signal could be further injected into the victim user's device to obtain legitimate access for further curious and malicious attacks. Specifically, *ActListener* first monitors the environmental noises and adds them into the converted signals to avoid user awareness and attack detection by a simple method. As mentioned in Section II-A, the signal processing in WiFi-based user identification usually eliminates the noise in the received signals, so the added noises would not degrade the attack performance. Then, the generated signals could be injected into the victim user's Rx via digital access, such as remote injection [41] or accessible firmware overwrite [42]. Even worse, if the adversary is under the WiFi coverage of the victim, he/she may build a spoofing AP to attack the linkage with the victim's legitimate Rx for on-site injection, i.e., transmitting a fake package with the converted signals for impersonation. The key challenge of such injection manners is to replay the converted signal with an extra transmitter or manipulate data inside the victim user's device, which can be implemented by existing wireless transmission and spoofing attacks. We omit their technical details since they are out of the scope of this paper.

Since *ActListener* recovers the input of the recognition and identification, i.e., the signal pattern received by legitimate Rx, the attack could be effective regardless of the machine learning models of the activity recognition and user identification, i.e., realizing a black-box attack.



(a) Office. (b) Apartment. (c) Lab.

Fig. 10. Floor maps of the three environments.

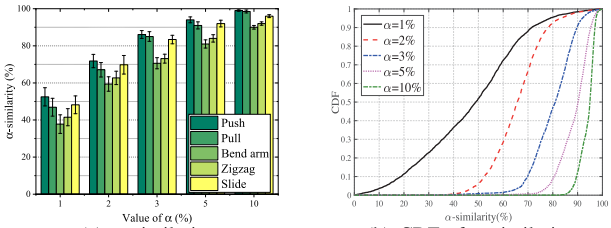
IV. Performance Evaluation

In this section, we evaluate the performance of *ActListener* under the collected data in three different environments.

A. Experimental Setup and Methodology

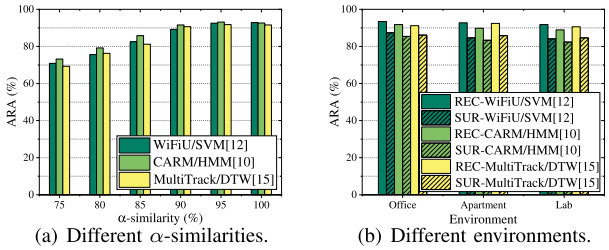
We implement *ActListener* on a laptop (i.e., HP Pavilion 14), integrating an Intel 5300 wireless NIC with three antennas, as the Sp. Another desktop (i.e., Dell E6430) is integrated with the same kind of NIC to act as the legitimate Rx. The desktop is responsible to train the activity recognition model, serving as a human-computer interaction terminal. Both devices are deployed with CSI tool [43]. Also, a wireless access point (i.e., TP-Link WDR5620), is employed as the Tx, which continuously transmits 802.11n signals for sensing.

We collect real-world CSI data for activity recognition and user authentication under the same experimental setup. Specifically, we recruit 15 volunteers as victim users, and select 5 widely-used activities for human-computer interactions, i.e., push, pull, bend arm, zigzag, and slide. The volunteers are with the ages of $[19, 43]$, heights of $[1.59, 1.8]m$, and weights of $[48, 74]kg$. We repeat the experiments in three real environments, i.e., an office, an apartment, and a lab. The three environments are of different sizes, i.e., $3.2m \times 2.8m$, and $4.1m \times 3m$ and $5.8m \times 4.2m$ respectively, and various furniture layouts, inducing different multipath effects during the sensing, whose floor maps are shown in Fig. 10. In each environment, we place Tx and Rx with a distance of $2m$ as the activity recognition system, between which each volunteer performs the defined behaviors in the midpoint of Tx-Rx connection for the recognition, which is almost consistent with that in most WiFi sensing researches [4], [5], [6], [7], [8], [9], [10], [23], [24], [44]. In each experiment, each victim user randomly performs a selected behavior for human-computer interactions or other daily activities. On the other hand, we place another WiFi receiver Sp as the compromised WiFi device at a distance of $1.5m$ away from Tx and directly facing the victim user (i.e., the angle between Tx-Rx and Tx-Sp connections is $60°$) to eavesdrop on the broadcasting WiFi signals, as the default setting. The adversary further employs the eavesdropped signals to query the desktop for evaluating the performance of activity surveillance and user impersonation. Each volunteer repeats each behavior 50 times in each environment, thus collecting 11,250 samples in total, which are divided into training and testing subsets by 8:2 for evaluation. The experiments on volunteers are validated by the Institutional Review Board (IRB) in Zhejiang University.



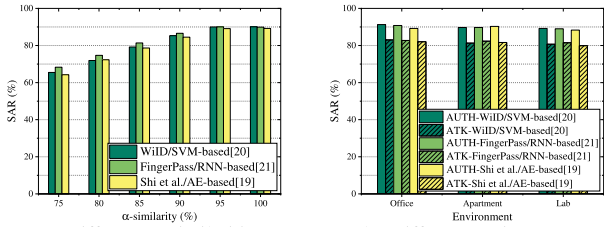
(a) α-similarity. (b) CDF of α-similarity.

Fig. 11. α-similarity of *ActListener* with different values of α.



(a) Different α-similarities. (b) Different environments.

Fig. 12. ARA of *ActListener* on different activity recognition models under different impacts.



(a) Different α-similarities. (b) Different environments.

Fig. 13. SAR of *ActListener* on different user identification models under different impacts.

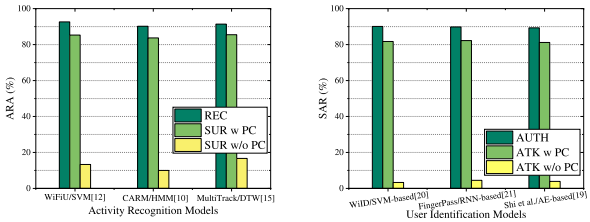
We define several metrics to evaluate the performance of the activity modeling-based signal conversion (as mentioned in Section III-C.2) and generative model-based signal calibration (as mentioned in Section III-C.3) methods.

- *α-similarity.* Assume the converted CSI is X and the ground truth is Y, whose normalized CSIs are x and y respectively. We define the sample error as $e_t = \frac{|x_t - y_t|}{y_t}$, where x_t and y_t are the CSI sample at time index t. Hence, the α-similarity is defined as the percentage of sample points satisfying $e_t \leq \alpha$ among all samples in a CSI.
- *Activity Recognition Accuracy (ARA).* The probability that an adversary uses a converted signal eavesdropped from a victim user performing an activity A is exactly authenticated as A by the activity recognition system.
- *Spoofing Acceptance Rate (SAR).* The probability that a spoofer or an adversary uses a converted signal eavesdropped from a victim user U is exactly authenticated as U by the targeted WiFi-based user identification.

B. Performance of Activity Modeling-Based Conversion

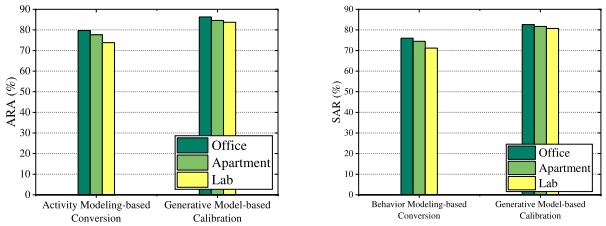
1) Performance on Signal Conversion: We first evaluate the similarity between converted signal by the behavior modeling conversion in *ActListener* and originally received signal. Fig. 11(a) shows the α-similarity with different α values under different gestures and volunteers. We find that the α-similarities under linear behaviors are better than those under non-linear behaviors. Also, the α-similarity increases with the increase of α value. This is because a larger value of α indicates higher error tolerance of the similarity metric. As the value of α approaches 5%, the α-similarity is above 85%, indicating a larger possibility of using converted signals for successful activity surveillance and impersonation. On the other hand, it can be also observed that the standard deviations of α-similarity under the same gesture and α are all below 5%. Specifically, when the value of α is 10, its standard deviations under different gestures are all within 1%. These results indicate that different users would not introduce significant differences into the converted signal, thus not interfering with the following activity surveillance and impersonation attacks. We also evaluate *ActListener*'s performance in terms of CDF of α-similarity under different α values, whose result is shown in Fig. 11. It can be observed that for 90% samples, the α-similarities between converted and originally received signals are above 20%, 50%, 70%, 80% and 90% under different values of α respectively. From the perspective of statistics, 5% is usually regarded as the boundary of confidence level. Along with this principle, we can find that the average α-similarity is 88.4%, and 90% samples are within an 80% α-similarity with $\alpha = 5\%$, indicating satisfactory performance of recovering the originally received signals from the eavesdropped ones.

2) Performance on Activity Surveillance: We further evaluate ARA of *ActListener* on various activity recognition models. We implement three representative machine learning models employed in existing WiFi activity recognition researches, i.e., Support Vector Machine (SVM) [8], Hidden Markov Model (HMM) [6], and Dynamic Time Warping (DTW) [23], for evaluation. Note that we feed the original signal patterns, instead of the adversary's converted ones, into the systems in the 100% α-similarity evaluation for the three activity recognition systems as baselines. Fig. 12(a) shows ARA of *ActListener* using signals with different α-similarities on different models. We can see that the ARA of using signals from activity modeling-based conversion in *ActListener* for the activity surveillance under different models increases as the α-similarity between signals increases. Specifically, as the α-similarity increases to 85%, the average ARA approaches 83.1%. Compared with the users' average ARA of 92.3%, the compromised Sp's ARA under 85% α-similarity only decreases within 10%. Moreover, as the result in Fig. 11 shows, 75% signals converted by the activity modeling-based conversion could achieve an 85% α-similarity with the originally received signals, indicating most eavesdropped signals by the compromised Sp could be used to achieve similar ARAs of activity recognition for surveillance. On the other hand, we also find the ARA under different models exhibits minute differences, demonstrating *ActListener* is a black-box attack without the requirement of model details. Fig. 12 shows ARA of *ActListener* in different environments on different models. We can see that in the three environments, the compromised Sp's ARAs (i.e., SUR in the figure) are all smaller than the user's ARAs (i.e., REC in the figure) within 10% only, which is consistent with the previous result. Also, the ARA in the three environments varies within 5%, indicating *ActListener* is robust to different environments.



(a) ARA. (b) SAR.

Fig. 14. ARA and SAR of *ActListener* with and without Pattern Conversion (PC) on different activity recognition and user identification models.



(a) ARA. (b) SAR.

Fig. 15. ARA and SAR of *ActListener* using different methods in different environments.

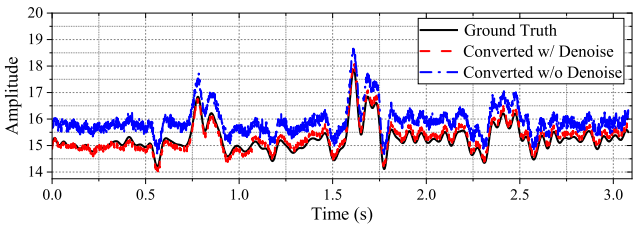


Fig. 16. Comparison among ground truth, converted signals of a "Slice" activity with and without generative model-based denoising.

3) Performance on Impersonation Attack: We also evaluate SAR of *ActListener* on various user identification models to demonstrate its threat for impersonation attacks. We implement three representative WiFi-based user identification models, i.e., Support Vector Machine (SVM) [18], Recurrent Neural Network (RNN) [12], and AutoEncoder (AE) [17], for evaluation. Following their original implementation, we feed the high-level hidden vectors of the training samples from 10 volunteers to an SVM classifier for user enrolment and those of the left 5 volunteers for spoofer rejection. Note that the experimental setup for 100% α-similarity evaluation is the same as the previous one. Fig. 13(a) shows SAR of *ActListener* using signals with different α-similarities on different models. We can see that the SAR of *ActListener* increases as the α-similarity between signals increases, which is consistent with the result on attacking activity recognition models. Compared with the users' average SAR of 89.7%, the adversary's SAR under 85% α-similarity decreases within 20%, which is larger than that on activity recognition, due to the less robustness of user identification. But with 85% α-similarity, *ActListener* still could achieve an average SAR approaching 80%. Moreover, we find the SAR on attacking different models exhibits minute differences, further demonstrating *ActListener* is a black-box attack without the requirement of model details regardless of specific tasks. Fig. 13 shows SAR of *ActListener* in different environments on attacking different models. We can see that in the three environments, the adversary's SARs (i.e., ATK in the figure) are all smaller than the user's SARs (i.e., AUTH in the figure) within 10% only, which is consistent with the corresponding result of activity recognition. Also, the SAR in the three environments varies within 5%, further demonstrating *ActListener* is robust to different environments.

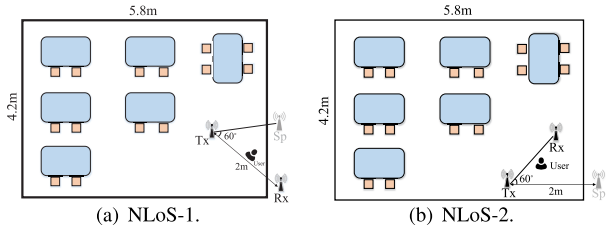
4) Effectiveness of Pattern Conversion: To further validate the effectiveness of our proposed pattern conversion approach in *ActListener*, we further evaluate the ARAs and SARs of *ActListener* with and without pattern conversion on different activity recognition and user identification models, as shown in Fig. 14. We can see that, compared to *ActListener* with pattern conversion, the ARA and SAR without pattern conversion dramatically decrease below 20% and 5% respectively. These results indicate that the pattern conversion plays a critical role in *ActListener* in activity surveillance and impersonation attacks.

C. Performance of Generative Model-Based Calibration

We also evaluate the performance of the generative model-based calibration in *ActListener*. Though the signal calibration is expected to obtain a more similar signal to the originally received one after signal conversion, its intrinsic principle heads to fit the activity recognition or spoof user identification only, instead of generating a similar signal. Toward this end, we only evaluate the ARA and SAR using the signal from the generative models for the activity surveillance or impersonation attack. Fig. 15(a) shows ARA of *ActListener* using different methods in different environments. It can be observed that the average ARA of generative model-based calibration is 7.9% larger than that of activity modeling-based conversion, indicating the improvement of introducing the generative model for noise-resistant signal calibration. Also, the standard deviation of ARA decreases from 3.0% to 1.3% after involving the generative model-based calibration.

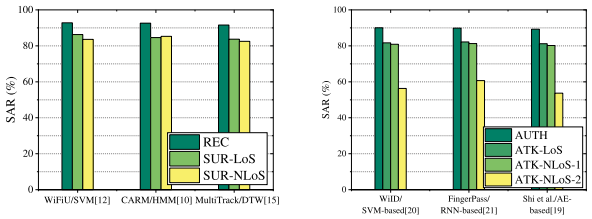
Fig. 15(b) shows corresponding SAR of *ActListener*. It can be observed that the average SAR of generative model-based calibration is 7.7% larger than that of behavior modeling-based conversion, which is similar to that of activity recognition. Also, the SAR standard deviation decreases 1.6% after involving the calibration. These results further demonstrate that the proposed generative model is efficient to resist noises in recovered CSI of WiFi signals for improving the robustness.

To further illustrate the improvement of our activity modeling-based conversion and generative model-based calibration in the recovered signals, we present legitimate signals (i.e., ground truth), as well as the converted and calibrated signals under "slice" for explicit comparison, as shown in Fig. 16. We can observe a similar time-aligned pattern between the legitimate and converted signals, indicating the effectiveness of signal conversion. However, there also exhibits an obvious difference in terms of the amplitude, due to the presence of channel noises. After applying signal calibration on the converted signal, we can see that the amplitude difference is greatly suppressed so that the legitimate and calibrated signals exhibit higher similarity, validating the necessity of *ActListener*'s calibration of noise resistance.



(a) NLoS-1. (b) NLoS-2.

Fig. 17. Floor maps of the two NLoS scenarios.



(a) ARA. (b) SAR.

Fig. 18. ARA and SAR of *ActListener* in LoS and NLoS scenarios.

D. Impact of NLoS Scenarios

Except for the Line-of-Sight (LoS) scenario, we also evaluate the performance of *ActListener* in the Non-LoS (NLoS) scenarios, whose floor map is shown in Fig. 17(a). Fig. 18(a) shows ARA of *ActListener* in NLoS scenarios. It can be observed that the ARAs in NLoS only exhibit a slight decrease compared with that in LoS, which is still over 80%. This is because though the signal diffracts and penetrates through the wall, the received signals by Sp and Rx are the ones directly interacted by the user body, leading to predictable multipath. Especially when WiFi operating bands increase from 2.4GHz to 5GHz (or even 60GHz), the penetration becomes more significant, expecting to achieve better performance. Such a result demonstrates that WiFi activity surveillance even induces more severe threats than vision-based ones.

We also evaluate the performance of *ActListener* on attacking user identification under NLoS scenarios. Considering the less robustness of user identification, we further refine the NLoS scenarios into two different ones, i.e., *NLoS-1*: both Rx and Sp are on one side of a wall, while Tx is on the other side, as shown in Fig. 17(a); *NLoS-2*: only Sp is behind a wall, while Rx and Tx are in the front, as shown in Fig. 17(b). Considering the usage scenarios of WiFi-based user identification, the volunteers performing gestures for login are always on the same side with Rx. Fig. 18(b) shows SAR of *ActListener* in NLoS scenarios. We can see that SAR in NLoS-1 is still over 80%, while that in NLoS-2 significantly decreases below 60%. This is because both Sp and Rx are on the same side of the wall in NLoS-1. Though the signal diffracts and penetrates through the wall, the received signals by Sp and Rx are the ones directly interacted by the user body. Hence, the introduced changes should remain the same for the two pairs, leading to predictable multipath. On the contrary, due to different placements of Sp and Rx in NLoS-2, the real-world propagation of signals probably involves complex diffraction and penetration, violating the models provided in Section III-C.2, and thus introduces significant performance degradation. These results suggest that *ActListener* is still effective in NLoS scenarios with a consistent setting between the Tx-Rx and Tx-Sp pairs.

TABLE I

ARA OF *ActListener* UNDER DIFFERENT DISTANCES ON DIFFERENT MODELS

	WiFiU/SVM[12]	CARM/HMM[10]	MultiTrack/ DTW[15]
REC	92.8%	92.6%	91.6%
SUR-1.5m	85.3%	83.7%	85.5%
SUR-1.6m	84.4%	82.8%	84.3%
SUR-1.8m	81.8%	80.2%	80.8%
SUR-2m	74.3%	73.1%	72%

TABLE II

SAR OF *ActListener* UNDER DIFFERENT DISTANCES ON DIFFERENT MODELS

	WiID/ SVM-based[20]	FingerPass/ RNN-based[21]	Shi et al./ AE-based[19]
AUTH	90.1%	89.8%	89.3%
ATK-1.5m	81.7%	82.2%	81.2%
ATK-1.6m	80.5%	81.0%	80.7%
ATK-1.8m	79.0%	78.7%	79.8%
ATK-2m	72.0%	71.5%	70.2%

E. Impact of Distances and Angles

We also evaluate the impact of distances and angles of Sp on the performance of *ActListener*. To be consistent with the settings in Section IV-A, the location is defined as the distance between Tx and the compromised Sp, and the angle between Tx-Rx and Tx-Sp connections. We repeat the aforementioned experiments under the distance of $[1.5m, 2m]$ and the angle of $[-60°, 60°]$. Tables I and II show ARA and SAR of *ActListener* under different distances on different models. We can observe that both ARA and SAR decrease as the increase of distance. This is because the intensity of WiFi signals degrades as the range between Tx and Rx/Sp increases, indicating a less reliable signal eavesdropped by the compromised Sp. But both ARA and SAR could still be larger than 80% within the distance of $1.8m$ and $1.6m$ on activity recognition and user identification, respectively.

On the other hand, Tables III and IV show ARA and SAR of *ActListener* under different angles on different models. It can be seen that except the angle directly facing the victim user (i.e., 60°), both ARA and SAR decrease in varying degrees. In particular, under the angle of $-60°$ and $-30°$, the ARAs both significantly decrease below 55% on average. And similar trend can be observed in SAR as shown in Table IV. This is because under these two angles, the compromised Sp is placed behind the victim user, thus hardly receiving the signals interacting with the victim user's moving body. But for other angles, the ARA and SAR could be above 80%, indicating the high possibility of successful activity surveillance under different angles.

V. DISCUSSIONS

In this section, we discuss some practical issues about *ActListener* and provide several countermeasures to defend such an attack.

Active Attack vs. Passive Attack. For a WiFi-based activity recognition system, an adversary could launch either an active attack (i.e., injecting interfering noises to disable the system), or a passive attack (i.e., the activity surveillance in this paper). It may not be difficult for an adversary to

TABLE III
ARA OF *ActListener* UNDER DIFFERENT ANGLES ON DIFFERENT MODELS

	WiFiU/SVM[12]	CARM/HMM[10]	MultiTrack/ DTW[15]
REC	92.8%	92.6%	91.6%
SUR-60°	86.8%	84%	84.8%
SUR-30°	86.4%	83.9%	85.3%
SUR-0°	85.3%	83.7%	85.5%
SUR-−30°	54.8%	49.5%	53.4%
SUR-−60°	55.8%	49.5%	53.4%

TABLE IV
SAR OF *ActListener* UNDER DIFFERENT ANGLES ON DIFFERENT MODELS

	WiID/ SVM-based[20]	FingerPass/ RNN-based[21]	Shi et al./ AE-based[19]
AUTH	90.1%	89.8%	89.3%
ATK-60°	84.0%	83.6%	82.5%
ATK-30°	82.1%	83.0%	81.8%
ATK-0°	81.7%	82.2%	81.2%
ATK-−30°	50.3%	48.9%	51.0%
ATK-−60°	45.6%	44.3%	48.9%

launch an active attack by exploring the noisy nature of CSI in state-of-the-art WiFi-based activity recognition systems. On the contrary, the passive attack is much more complex than the active one, and also more severe in practice due to its imperceptibility. Our work overcomes the challenges and demonstrates the feasibility of activity surveillance on WiFi-based activity recognition systems.

Untargeted Attack vs. Impersonation Attack. To attack a WiFi-based user identification, an adversary could launch either an untargeted attack to disable the system, or a targeted attack (i.e., an impersonation attack) to bypass the access control. It may not be difficult for an adversary to launch an untargeted attack by exploring the noisy nature of CSI in state-of-the-art WiFi-based user identifications. On the contrary, the impersonation attack is much more complex than the untargeted one, and also more severe in practice due to its imperceptibility. Our work overcomes the challenges and demonstrates the feasibility of the impersonation attack on WiFi-based user identifications.

Advanced Characteristic vs. Attack Success. As mentioned in the related work, most of existing WiFi-based activity recognition and user identification systems are domain-specific and thus sensitive to variations of environment, orientation, location, etc. Hence, direct replaying the eavesdropped WiFi signal without the pattern conversion of *ActListener* to the target system would significantly downgrade the attack success. Latest researches [45], [46] have enabled WiFi-based activity recognition in cross-domain scenarios for better user experience and flexibility. However, such an advanced characteristic probably introduces more severe vulnerability into the systems. Theoretically, a cross-domain WiFi-based activity recognition treats signal patterns received from various locations, orientations, or environments as the same one for a distinct user, i.e., the system would recognize any signals received around the victim user as the same activity pattern, even from a different position. This indicates the attack efforts may even be released for an adversary when attacking a cross-domain WiFi-based activity recognition. But in practice, the cross-domain WiFi-based systems still yield poor performance, due to the significant difference among various domains. In this case, *ActListener* could also be applied to improve the attack success. Therefore, more security concerns should be addressed before the practical deployments of WiFi-based activity recognition.

Extension on Emerging WiFi Standard. Compared to the omni-directional WiFi 4 (802.11n) signal used in the current attack, the advanced WiFi 6 (802.11ax) introduces beamforming techniques to enable directional downlink communication for improving the data rate. Such an advanced characteristic should not downgrade the attack performance of *ActListener*. This is because the beamforming technique is actually realized by fusing the multiple signals received from different antennas, which indicates the signal received from each antenna still propagates omni-directionally. Hence, when implementing *ActListener* on WiFi 6, we can realize the signal recovery based on the signal on each antenna, instead of using that after signal beamforming.

Countermeasures. To defend *ActListener*, the countermeasures could be applied on different steps of *ActListener*. The key of *ActListener* is to first eavesdrop on a user's behaviors from the omni-directional signals, and then employ it for activity surveillance by querying. Hence, the most straightforward countermeasure is to prevent your devices (i.e., the Rx) from signal injection by the firewall setup or physical protection. This is appropriate for some public environments (such as an office), where the user could not prevent his/her behavioral patterns from potential leakage in the air easily. Another way is to protect the connection to users' own router (i.e., the Tx) carefully, by setting an isolated network (such as a guest network) or simply rejecting any external connection. This scenario is more suitable for private spaces, such as a house or apartment. In addition, our evaluations actually reveal one potential solution by placing the Rx in specific locations. As demonstrated in Section IV-D, different LoS settings between Tx-Rx and Tx-Sp pairs could downgrade the surveillance performance. Based on this observation, the users can place all adjacent Rxes with different Tx-Rx blockages, i.e., if one Rx is placed in LoS with Tx, its adjacent Rxes should be placed in NLoS with Tx. Such a straightforward approach serves as a potential countermeasure to resist the attack.

VI. RELATED WORK

In this section, we discuss the key researches about WiFi sensing, as well as the attacks on wireless systems.

WiFi Sensing. In 1997, WiFi was invented and first released for consumers, and become pervasive in indoor environments nowadays. Because of its wide existence, researchers realize the great potential of using WiFi to implement low-cost and device-free sensing applications. Enormous researches employ WiFi signals to implement various applications, such as vital sign detection [47], activity recognition [4], [6], indoor localization [48], finger gesture tracking and recognition [10]. Among them, activity and gesture recognition attracts more attention to realize privacy-preserving natural human-computer interaction methods. Early work [4] demonstrates the feasibility of using WiFi to recognize human daily activities, inspired by which, following works [5], [6], [7], [8], [9], [10]

explore the sensing capability of WiFi on more fine-grained activity and gesture recognition. Recent studies [23], [24] even investigate the feasibility of WiFi sensing in multi-subject scenarios. Latest work [44] even demonstrates realizing an imaging system based on WiFi signals, further extending the usage of WiFi sensing.

With the rapid development of WiFi sensing, users' privacy concern also increases due to the stored sensitive information facilitating intelligent services. Hence, recent research efforts [12], [13], [14], [15], [16] further push WiFi sensing to individual identification area for rigorous access control, extending the usage of WiFi sensing. Early studies [25], [26] sense user walking gait using WiFi to distinguish different individuals. Following works [16], [17], [18] release sensing behaviors from gait to various common human activities for user identification. Except for coarse-grained activities, some researches [12], [13], [14], [15] even realize the fine-grained finger gesture sensing for user identification, and extend them to gesture-free and multi-person scenarios.

Attacks on Wireless Systems. Due to the omni-directional broadcasting manner of wireless signals, wireless systems are vulnerable to various attacks by nature. Generally, wireless-oriented attacks are categorized into the insider attack and outsider attack [49]. The insider attack requires compromising the wireless access point by injecting malicious software or accessing the device physically in advance, which is rarely realized in practice. Instead, the outsider attack only utilizes the omni-directional broadcasting signals for eavesdropping [50], [51], [52], [53], [54], [55] or disabling [49] purposes. Although many researchers realize the vital importance of providing security guarantees for wireless systems, this problem remains an open issue and has not been fully addressed yet [56]. Some works propose anti-eavesdropping systems in terms of key establishment [51], message concealing [50], signal leakage detection [53], etc., but the requirement of modifying commercial systems constrains its wide deployment. Following work [54] designs a unique signal transmitting scheme, relying on commercial single-antenna device only, to detect the malicious eavesdropper. However, all the aforementioned works are targeted at wireless communication, without the investigation of gradually developed wireless sensing. More recent studies start to explore eavesdropping attacks on WiFi sensing for password and keystroke inference [57], [58] or introduce adversarial example attacks [59], [60], [61] to undermine WiFi-based gesture or behavior recognition systems, but the feasibility of exploiting WiFi sensing for activity surveillance and user impersonation remains unexplored.

Different from them, our work aims to reveal the privacy leakage threat where an arbitrary WiFi infrastructure could turn to surveillance or impersonation equipment, which requires no direct physical access to legitimate users' devices and prior knowledge of user behaviors, model details and device locations.

VII. Conclusion

In this paper, we demonstrate *ActListener*, which could employ a compromised WiFi infrastructure in any location to realize the activity surveillance and impersonation. In particular, we first model the CSI of propagating WiFi signals induced by victim user's behaviors, and design the convert function to recover the legitimate signals from the adversary's received signals, so that the signal from a WiFi infrastructure in any location could be used for activity surveillance and impersonation. After that, we further design a generative model-based neural network to calibrate the converted signal for resisting the always-existed noises in CSI of WiFi signals. Experimental results demonstrate that *ActListener* can achieve good performance on recovering legitimate signals to retrieve user activities semantics and user identity.

References

[1] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, p. 46, 2019.

[2] O. C. News. (2020). *U.S. Households Will Have an Average of 20 Connected Devices by 2025*. [Online]. Available: https://www.cordcuttersnews.com/us-households-will-have-an-average-of-20-connected-devices-by-2025/

[3] P. I. Technology. *Perspicace Intelligent Technology—AI Creates Happy Life*. [Online]. Available: https://www.perspicace-china.com

[4] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. ACM MobiCom*, Maui, HI, USA, 2014, pp. 617–628.

[5] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. ACM MobiCom*, Sep. 2015, pp. 90–102.

[6] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. ACM MobiCom*, New York, USA, 2015, pp. 65–76.

[7] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1472–1480.

[8] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM UbiComp*, Sep. 2016, pp. 363–373.

[9] O. Zhang and K. Srinivasan, "Mudra: User-friendly fine-grained gesture recognition using WiFi signals," in *Proc. ACM CoNEXT*, Dec. 2016, pp. 83–96.

[10] S. Tan and J. Yang, "WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proc. ACM MobiHoc*, Paderborn, Germany, 2016, pp. 201–210.

[11] K. Kim, L. Boelling, S. Haesler, J. N. Bailenson, G. Bruder, and G. F. Welch, "Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, D. Chu, J. L. Gabbard, J. Grubert, and H. Regenbrecht, Eds. Munich, Germany: IEEE, Oct. 2018, pp. 105–114.

[12] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi," in *Proc. ACM MobiHoc*, Catania, Italy, Jul. 2019, pp. 201–210.

[13] H. Kong, L. Lu, J. Yu, Y. Chen, and F. Tang, "Continuous authentication through finger gesture interaction for smart homes using WiFi," *IEEE Trans. Mobile Comput.*, vol. 20, no. 11, pp. 3148–3162, Nov. 2021.

[14] H. Kong et al., "MultiAuth: Enable multi-user authentication with single commodity WiFi device," in *Proc. ACM MobiHoc*, Shanghai, China, Jul. 2021, pp. 31–40.

[15] H. Kong et al., "Push the limit of WiFi-based user authentication towards undefined gestures," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, London, U.K., May 2022, pp. 410–419.

[16] H. Kong, L. Lu, J. Yu, Y. Chen, X. Xu, and F. Lyu, "Toward multi-user authentication using WiFi signals," *IEEE/ACM Trans. Netw.*, vol. 31, no. 5, pp. 2117–2132, 2023.

[17] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in *Proc. ACM MobiHoc*, Chennai, India, 2017, pp. 1–10.

[18] M. Shahzad and S. Zhang, "Augmenting user identification with WiFi based gesture recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, p. 134, 2018.

[19] IEEE. (2020). *IEEE P802.11—WLAN Sensing Study Group*. [Online]. Available: https://www.ieee802.org/11/Reports/senstig_update.htm

[20] E. Au, "New standards initiative for using Wi-Fi for sensing [Standards]," *IEEE Veh. Technol. Mag.*, vol. 15, no. 1, p. 119, Mar. 2020.

[21] L. Lu, Z. Ba, F. Lin, J. Han, and K. Ren, "ActListener: Imperceptible activity surveillance by pervasive wireless infrastructures," in *Proc. IEEE 42nd Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Bologna, Italy, Jul. 2022, pp. 776–786.

[22] P. Security. (2018). *Thousands of Home Routers Hacked—What Can You do?* https://www.pandasecurity.com/en/mediacenter/mobile-news/routers-hacked/

[23] S. Tan, L. Zhang, Z. Wang, and J. Yang, "MultiTrack: Multi-user tracking and activity recognition using commodity WiFi," in *Proc. ACM CHI*, May 2019, pp. 1–12.

[24] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. ACM MobiSys*, Jun. 2018, pp. 401–413.

[25] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Vienna, Austria, Apr. 2016, pp. 1–12.

[26] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "WiFi-ID: Human identification using WiFi signal," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, Washington, DC, USA, May 2016, pp. 75–82.

[27] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[28] B. News. (2019). *A Data Leak Exposed the Personal Information of Over 3,000 Ring Users*. [Online]. Available: https://www.buzzfeednews.com/article/carolinehaskins1/data-leak-exposes-personal-data-over-3000-ring-camera-users

[29] S. Media. (2021). *Camera Tricks: Privacy Concerns Raised After Massive Surveillance Cam Breach*. [Online]. Available: https://www.scmagazine.com/home/security-news/iot/camera-tricks-privacy-concerns-raised-after-massive-surveillance-cam-breach/

[30] F. Zhang et al., "Towards a diffraction-based sensing approach on human activity recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 1, p. 33, 2019.

[31] M. H. Weik, "Inverse square law," *Computer Science and Communications Dictionary*, 2000, p. 834.

[32] Y. Zhuo, H. Zhu, and H. Xue, "Identifying a new non-linear CSI phase measurement error with commodity WiFi devices," in *Proc. IEEE 22nd Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Wuhan, China, Dec. 2016, pp. 72–79.

[33] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM Conf. Special Interest Group Data Commun.*, London, U.K., Aug. 2015, pp. 269–282.

[34] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. USENIX Symp. Netw. Syst. Design Implement.*, Boston, MA, USA, 2013, pp. 71–84.

[35] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[36] H. Xue, J. Yu, Y. Zhu, L. Lu, S. Qian, and M. Li, "WiZoom: Accurate multipath profiling using commodity WiFi devices with limited bandwidth," in *Proc. 16th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Boston, MA, USA, Jun. 2019, pp. 1–9.

[37] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 3, pp. 328–339, 1989.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[40] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, Atlanta, GA, USA, 2013, pp. 1–16.

[41] C. Blancher. (2005). *Attacking WiFi Networks With Traffic Injection*. [Online]. Available: http://axellec.chez.com/securite/LSM2005/WirelessInjection_CedricBlancher_LSM2005_slides.pdf

[42] OpenWrt. (2020). *Flashing OpenWrt With WiFi Enabled on First Boot*. [Online]. Available: https://openwrt.org/docs/guide-user/installation/flashing_openwrt_with_wifi_enabled_on_first_boot

[43] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, Aug. 2010.

[44] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5451–5460.

[45] W. Jiang et al., "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, New Delhi, India, 2018, pp. 289–304.

[46] Y. Zheng et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst. Appl. Services*, Seoul, Republic of Korea, 2019, pp. 313–325.

[47] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. ACM MobiHoc*, Hangzhou, China, 2015, pp. 267–276.

[48] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.

[49] E. Shi and A. Perrig, "Designing secure sensor networks," *IEEE Wireless Commun.*, vol. 11, no. 6, pp. 38–43, Dec. 2004.

[50] S. Fang, T. Wang, Y. Liu, S. Zhao, and Z. Lu, "Entrapment for wireless eavesdroppers," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Paris, France, Apr. 2019, pp. 2530–2538.

[51] S. Fang, I. Markwood, and Y. Liu, "Manipulatable wireless key establishment," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Las Vegas, NV, USA, Oct. 2017, pp. 1–9.

[52] Y.-C. Tung, S. Han, D. Chen, and K. G. Shin, "Vulnerability and protection of channel state information in multiuser MIMO networks," in *Proc. ACM CCS*, Scottsdale, AZ, USA, 2014, pp. 775–786.

[53] A. Chaman, J. Wang, J. Sun, H. Hassanieh, and R. Roy Choudhury, "Ghostbuster: Detecting the presence of hidden eavesdroppers," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, New Delhi, India, Oct. 2018, pp. 337–351.

[54] T. J. Pierson, T. Peters, R. Peterson, and D. Kotz, "Proximity detection with single-antenna IoT devices," in *Proc. ACM MobiCom*, Los Cabos, Mexico, 2019, p. 21:1–21:15.

[55] M. Alyami, I. Alharbi, C. Zou, Y. Solihin, and K. Ackerman, "WiFi-based IoT devices profiling attack based on eavesdropping of encrypted WiFi traffic," in *Proc. IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2022, pp. 385–392.

[56] M. Fomichev, F. Álvarez, D. Steinmetzer, P. Gardner-Stephen, and M. Hollick, "Survey and systematization of secure device pairing," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 517–550, 1st Quart., 2017.

[57] J. Zhang et al., "Privacy leakage in mobile sensing: Your unlock passwords can be leaked through wireless hotspot functionality," *Mobile Inf. Syst.*, vol. 2016, Apr. 2016, Art. no. 8793025.

[58] Y. Meng, J. Li, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "Revealing your mobile password via WiFi signals: Attacks and countermeasures," *IEEE Trans. Mobile Comput.*, vol. 19, no. 2, pp. 432–449, Feb. 2020.

[59] Y. Zhou, H. Chen, C. Huang, and Q. Zhang, "WiADv: Practical and robust adversarial attack against WiFi-based gesture recognition system," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, p. 92, 2022.

[60] J. Liu, Y. He, C. Xiao, J. Han, L. Cheng, and K. Ren, "Physical-world attack towards WiFi-based behavior recognition," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, London, U.K., Jun. 2022, pp. 400–409.

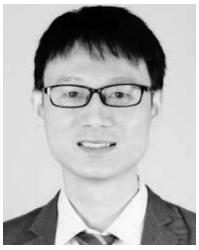
[61] L. Xu, X. Zheng, X. Li, Y. Zhang, L. Liu, and H. Ma, "WiCAM: Imperceptible adversarial attack on deep learning based WiFi sensing," in *Proc. 19th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Stockholm, Sweden, Sep. 2022, pp. 10–18.

Li Lu (Member, IEEE) received the B.E. degree from Xi'an Jiaotong University and the Ph.D. degree from Shanghai Jiao Tong University. He is currently a tenure-track Research Professor with the School of Cyber Science and Technology and the College of Computer Science and Technology, Zhejiang University. He also visited the Wireless Information Network Laboratory (WINLAB) and the Department of Electrical and Computer Engineering, Rutgers University. His research interests include the IoT security, intelligent voice security, mobile sensing, and ubiquitous computing. He was a recipient of the ACM China SIGAPP Chapter Rising Star Award, the ACM China SIGAPP Chapter Doctoral Dissertation Award, the Best Poster Runner-Up Award from ACM MobiCom 2022, and the First Runner-Up Poster Award from ACM MobiCom 2019.

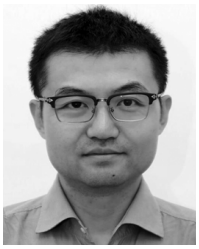
Meng Chen (Graduate Student Member, IEEE) received the B.E. degree in software engineering from Zhejiang University, where he is currently pursuing the Ph.D. degree with the School of Cyber Science and Technology. His research interests include mobile computing and AI security. He was a recipient of the Best Poster Runner-Up Award from ACM MobiCom 2022 and the Student Travel Grant of IEEE INFOCOM 2022.

Jiadi Yu (Senior Member, IEEE) received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2007. He is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Prior to joining Shanghai Jiao Tong University, he was with the Stevens Institute of Technology, USA, as a Post-Doctoral Researcher. He has published more than 100 refereed papers in international leading journals and key conferences in the areas of wireless communications and networking, mobile computing, and security and privacy. His current research interests include mobile computing and sensing, cyber security and privacy, the Internet of Things (IoT), and smart healthcare. He is a Senior Member of the IEEE Communication Society.

Zhongjie Ba received the Ph.D. degree in computer science and engineering from The State University of New York at Buffalo in 2019. He was a Post-Doctoral Researcher with the School of Computer Science, McGill University. He is currently a ZJU100 Young Professor with the College of Computer Science and Technology and the Institute of Cyberspace Research (ICSR), Zhejiang University, Hangzhou, China. His current research interests include the security and privacy aspects of Internet of Things, artificial intelligence-powered mobile sensing, and forensic analysis of multimedia contents.



Feng Lin (Senior Member, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering, Tennessee Tech University, Cookeville, TN, USA, in 2015. He was an Assistant Professor with the University of Colorado Denver, Denver, CO, USA; a Research Scientist with The State University of New York (SUNY) at Buffalo, Buffalo, NY, USA; and an Engineer with Alcatel-Lucent (currently, Nokia). He is currently a Professor with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University, China. His current research interests include mobile sensing, the Internet of Things security, biometrics, AI security, and the IoT applications. He was a recipient of the Best Paper Awards from ACM MobiSys'20, IEEE Globecom'19, and IEEE BHI'17. He was a recipient of the Best Demo Award from ACM HotMobile'18 and the First Prize Design Award from the 2016 International 3D Printing Competition.



Jinsong Han (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology. He is currently a Professor with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University. His work focuses on the IoT, smart sensing, mobile computing, and AI. He was the Winner of Hong Kong ICT Awards—Best Innovation and Research Award (Silver Award). He received the Best Paper Awards of the 2019 IEEE INFOCOM and the 2019 GLOBECOM and the 2018 ACM Xi'an Best Supervisor Award. He was selected as an Excellent Teachers of Computer Major in Chinese Colleges and Universities. He is a Senior Member of ACM.

Yanmin Zhu (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, in 2007. He is currently a Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Before that, he was a Research Associate with the Department of Computing, Imperial College London. His research interests include crowd sensing, big data analytics and systems, and cloud computing.

Kui Ren (Fellow, IEEE) received the Ph.D. degree from Worcester Polytechnic Institute, Worcester, MA, USA. He is currently the Dean and a Professor with the School of Cyber Science and Technology, Zhejiang University. He has published extensively in peer-reviewed journals and conferences. He is a Fellow of ACM, a Distinguished Lecturer of IEEE, and a past Board Member of the Internet Privacy Task Force, State of Illinois. He received several best paper awards, including IEEE ICDCS 2017, IWQoS 2017, and ICNP 2011. He received the NSF CAREER Award in 2011, the Sigma Xi/IIT Research Excellence Award in 2012, the UB SEAS Senior Researcher of the Year Award in 2015, the UB Exceptional Scholar Award for Sustained Achievement in 2016, and the IEEE CISTC Technical Recognition Award in 2017. He currently serves on the editorial boards for IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON SERVICE COMPUTING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE WIRELESS COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, and *SpringerBriefs on Cyber Security Systems and Networks*.