A Behavior Privacy Preserving Method towards RF Sensing

Jianwei Liu[1], Chaowei Xiao[2,8], Kaiyan Cui[3,4], Jinsong Han[1]✉, Xian Xu[1], Kui Ren[1,5,6], and Xufei Mao[7]

[1]Zhejiang University, China
[2]Arizona State University, USA
[3]Xi'an Jiaotong University, China
[4]The Hong Kong Polytechnic University, China
[5]Key Laboratory of Blockchain and Cyberspace Governance of Zhejiang Province, China
[6]Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, China
[7]Dongguan University of Technology, China
[8]Nvidia Research

***Abstract*—Recent years have witnessed the booming development of RF sensing, which supports both identity authentication and behavior recognition by analysing the signal distortion caused by human body. In particular, RF-based identity authentication is more attractive to researchers, because it can capture the unique biological characteristics of users. However, the openness of wireless transmission raises privacy concerns since human behaviors can expose the massive private information of users, which impedes the real-world implementation of RF-based user authentication applications. Unfortunately, it is difficult to filter out the behavior information from the collected RF signals.**

In this paper, we propose a privacy-preserving deep neural network named *BPCloak* to erase the behavior information in RF signals while retaining the ability of user authentication. We conduct extensive experiments over mainstream RF signals collected from three real wireless systems, including the WiFi, Radio Frequency IDentification (RFID), and millimeter-wave (mmWave) systems. The experimental results show that *BPCloak* significantly reduces the behavior recognition accuracy, *i.e.*, 85%+, 75%+, and 65%+ reduction for WiFi, RFID, and mmWave systems respectively, merely with a slight penalty of accuracy decrease when using these three systems for user authentication, *i.e.*, 1%-, 3%-, and 5%-, respectively.

I. Introduction

Recently, RF sensing has been receiving widespread attention, which utilizes RF signals to collect various object information. Compared with related solutions (*e.g.*, camera- and wearable-based solutions), it not only supports non-line-of-sight scenarios, but also does not require users to carry any devices. More importantly, RF sensing can mitigate visual privacy concerns since it would not reveal human's visual information. Therefore, non-intrusive and contactless RF sensing enables a variety of essential human-computer interaction applications, such as user identity authentication [1], behavior recognition [2], and tracking [3].

Among these applications, user authentication is a particular interest of researchers in recent years [1], [4]–[6]. This is because RF signals can penetrate human bodies and carry biological or physiological characteristics [4]. By analyzing the received RF signals, one can extract user-dependent and unique identity information to perform authentication. Besides, the identity information in the RF signals is difficult to be stolen/duplicated/forged, compared with other kinds of identity information (*e.g.*, fingerprint and facial features) [7].

However, RF-based user authentication also introduces privacy concerns, because RF signals not only can record users' identity information, but also are able to sense users' behavior information. For example, while users may enjoy RF-based user authentication [4] provided by a service provider (SP) [8], it also allows the SP to steal their behavior privacy. A malicious SP could utilize the received signals to mine users' private behaviors, such as body movements [9], arm motions [2], and finger traces [10]. The SP can even speculate the users' personality [11], psychology [12], and some private passwords [13] according to the behavior information.

Given these severe privacy and security consequences, one question is naturally raised: *can the RF signal only record the identity information while avoiding capturing the behavior one?* To answer this question, we analyze the human influence on the RF signal. The analysis results show that as long as a person appears within the signal sensing range for identity authentication, his identity information and behavior information will be inevitably recorded at the same time.

In this paper, we aim to propose a 'once-deploy-forever-use' method to protect the behavior information in RF signals. This method enables users to 'disable the behavior recognition' by filtering their private behavior information out as much as possible from RF signals, yet still supports RF-based user authentication with a high accuracy.

However, achieving this goal is difficult due to the following challenges. First, the most intuitive way to protect behavior privacy is to separate identity information and behavior information, and then only retain the former. However, separating the behavior and identity information directly is infeasible, because RF signals are not as intuitive as visual

978-0-7381-3207-5/21/$31.00 ©2021 IEEE

data (*e.g.*, image/video). Meanwhile, the exact function relationship between these two kinds of information is agnostic (detailed in Section III-A). Second, it is not acceptable to degrade/ruin/erase the identity information too much when excluding the behavior information.

To overcome the first challenge, we transform the intractable separation problem into a simple similarity problem. We first dig out the reason that RF signals can be used for identity/behavior recognition. We find that signals are spatially distinguishable for both identity and behavior. Intrinsically, erasing the behavior information is equivalent to erase the distinguishability among the signals corresponding to different behaviors. One way to achieve this goal is to increase the similarity among the signals of different behaviors. Therefore, we design a *Siamese network*-based [14] deep neural network, named *BPCloak*, to adjust the similarity to erase behavior information. To address the second challenge, we do not increase the similarity among the signals of different identities to maintain the distinguishability among different identities. Besides, we further introduce an identity classifier *BPCloak* to improve its identity feature extraction ability.

Specifically, we treat the signals corresponding to different behaviors with the same identity as similar, and the signals corresponding to the same behavior with different identities as dissimilar in the *BPClack*. In this way, *BPCloak* can erase the behavior information without destroying the identity information. To further improve the identity feature extraction ability of *BPCloak*, we leverage an identity classifier to facilitate the optimization of *BPCloak*. For the real-world deployment, users first collect a batch of signal samples (each piece of data in RF signals is termed as a signal sample) and generate a training set based on our training set construction method. After training *BPCloak* using the training set as well as a reasonably-designed loss function, *BPCloak* can erase behavior privacy. *BPCloak* takes signal samples as input and outputs behavior-irrelevant identity feature vectors. Since such a feature vector does not contain behavior information but contains sufficient identity information, it can be directly made public or uploaded to the SP to achieve accurate identity authentication without revealing behavior privacy.

We conducted comprehensive experiments on different RF systems including a commercial off-the-shelf (COTS) WiFi system, a COTS RFID system, and a COTS mmWave system to represent omnidirectional RF technique, backscatter RF technique, and directional RF technique, respectively. The experiment results on the WiFi/RFID/mmWave system demonstrated that *BPCloak* can decrease the behavior recognition accuracy from 99%/95%/99% to 11%/18%/28%, while the accuracy of identity authentication only dropped 1%/3%/5%.

This paper makes the following contributions:

- We observe the privacy issues in RF-based user authentication applications that lead to user's behavior information leakage.
- We propose a novel method, the core of which is *BPCloak*, to erase behavior privacy in RF signals without hurting the performance of RF-based user authentication.
- We conduct comprehensive experiments on three representative RF systems. The experiment results demonstrate that *BPCloak* can effectively erase the behavior information in the signal while retaining the high accuracy of identity authentication.

II. Privacy Concern in RF Signals

To answer the aforementioned naturally-raised question, we analyze the procedure of collecting RF signals and show the tight combination between the identity and behavior information in the collected signals. We then present a threat model and show how an adversary can compromise behavior privacy via *RF signals that have not been processed for behavior privacy preserving* (termed as vanilla RF signals).

A. RF Signal Collection

While collecting RF signals, the influence of a human body on the RF signals can be divided into static influence and dynamic influence according to whether the human body is static [4], [10] or dynamic [2], [15]. Both influences will inevitably embed identity information and behavior information in the signals. These two kinds of information lead to the distinguishability of signals in identity and behavior. Such distinguishability enables RF-based identity authentication and behavior recognition to be realized by means of feature vector extraction and machine learning classifiers [6], [16]. This distinguishability can be reflected by the distinguishability of signal indicators, such as amplitude (higher amplitude means higher signal energy) and phase. Below, we take the distinguishability of amplitude as an example for illustration.

In the static case, *i.e.*, the user keeps still, the identity is distinguishable in terms of the body's biomaterial and shape. As shown in Fig. 1(a) and (b), when two different persons pose the same static gesture, the received signals are identity-distinguishable because 1) the biomaterials of different persons are distinguishable [1] and the RF signals are sensitive to the material they pass through [17]; 2) the shapes of different persons are also distinguishable [18] and different shapes would cause different signal propagation paths by reflection [4], [19]. When the same signal experiences different persons

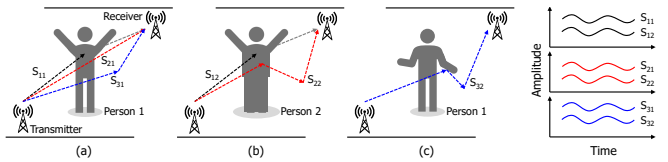


Fig. 1. The identity and behavior information are captured when the person is static.

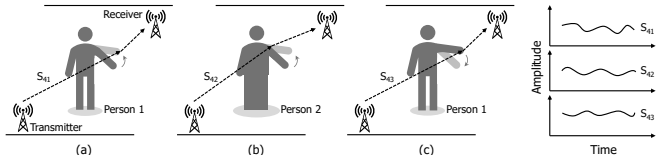


Fig. 2. The identity and behavior information are captured when the person is dynamic.

(see S_{11} and S_{12} or S_{21} and S_{22} in Fig. 1) in different scenarios, its received signal amplitudes are distinguishable. As shown in Fig. 1(a) and (c), when the same person poses different gestures, the received signals (S_{31} and S_{32}) are gesture-distinguishable because different gestures also cause different amplitudes [10].

In the dynamic case, the user is performing certain activities. Besides the aforementioned biomaterial and shape, we involve behavioral pattern (which is referred to as behavioral feature [7]). For two persons who perform the same activity, not only their biomaterials and body shapes will produce amplitude distinguishability, but their behavioral patterns of performing activities are also different from person to person [18], [20] (see S_{41} and S_{42} in Fig. 2). Similarly, when the same person performs two different activities (see S_{41} and S_{43}), different activities cause different signal propagation paths [21] and further produce distinguishability on amplitude variations. Therefore, identity information and behavior information are also recorded by signals in the dynamic case. From the above analysis, we conclude that as long as a person appears in the sensing range of the signal, his identity and behavior information will inevitably and simultaneously be captured. As an application, WiHF [22] have confirmed the feasibility of using the same RF signal simultaneously for user identification and gesture recognition.

B. Threat Model and Attack Test

Threat model: Once the vanilla RF signals are obtained by an adversary, the risk of behavior privacy leakage appears. We assume the adversary knows the feature extraction method of the user. The extracted features are used to identify users in user authentication applications. This assumption is reasonable because the category (*e.g.*, statistical scalar or original signal values) of the feature can be inferred according to the feature form.

Attack test: After the vanilla signals are acquired by the adversary, the adversary can leverage some techniques (*e.g.*, machine learning classifier) to speculate the behaviors related to the signals. To show the hazardness of such attacks, we conduct an attack experiment. We first ask a volunteer (victim) to performs six activities introduced in [23] in an environment to collect a batch of signal samples. Then we invite another volunteer (adversary) to collect another batch of signal samples in another environment in the same way, After extracting frequency features [23] from the two batches, we train a logistic regression (LR) classifier [24] with the adversary's features and test the classification accuracy with the victim's features. As a result, the activity classification accuracy is 64%+. This result demonstrates the feasibility of using machine learning classifiers to mine the behavior privacy from vanilla signals. Moreover, Zheng *et al.* [2] show that when the adversary extracts domain-independent features from vanilla signals, the malicious behavior classification accuracy can reach 92.7%, even if the adversary's signals and the victim's signals are collected by different persons in different environments.

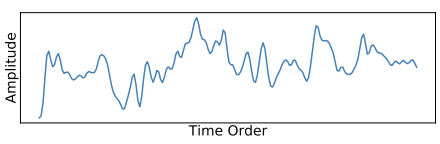


Fig. 3. The signal is not as intuitive as image/video.

III. Paving The Way For Privacy Preserving

In this section, we first introduce the reason why we use the deep learning technique to achieve behavior privacy preserving. Then, we transform the distinguishability problem into a similarity problem, which enables us to leverage *Siamese network* to change the similarity among signal samples.

A. Why Deep Learning?

To protect behavior privacy in vanilla signals, intuitive methods are to directly separate them via visual observation or numerical analysis (*e.g.*, parameters estimation [25]). However, these methods are infeasible due to the following reasons: 1) RF signal is not as intuitive as image/video. As shown in Fig. 3, one cannot figure out which part of the signal represents the identity or behavior information. 2) To separate the two kinds of information through numerical analysis, we need to build a function relationship between the two kinds of information. According to [26], such function relationship can be formulated as:

$$H = \sum_{k=1}^{P} (a_k^i + a_k^b) e^{-j2\pi f(\tau_k^i + \tau_k^b)}, \tag{1}$$

where H is the channel state information that describes how a signal propagates from transmitters to receivers. The P is the number of multi-paths. The a_k^i and τ_k^i are identity information-related amplitude and propagation delay, respectively. The a_k^b and τ_k^b are behavior information-related amplitude and propagation delay, respectively. Since P is unknown in reality, the exact function relationship between the two kinds of information is unknown, not mention to separate them through numerical analysis. Fortunately, we notice that a deep neural network can be trained to fit any function. Thus, it is potential to leverage the deep neural network to learn the complex function relationship between the two kinds of information and further separate them.

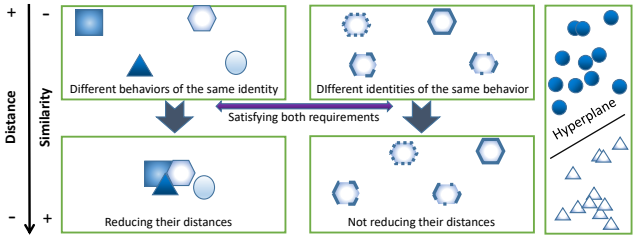


Fig. 4. Reducing the distances among the signals of different behaviors but the same identity while not reducing the distances among the signals of different identities but the same behavior.

B. From Distinguishability To Similarity

As mentioned in II-A, existing identity/behavior recognition techniques dominantly leverage the distinguishability among signals, extracted feature vectors, and classifiers to achieve high accuracy. Intrinsically, removing the behavior information in RF signals is to make the feature vectors extracted from the signal samples associated with different behaviors no longer be distinguishable in behavior. To this end, as shown in the left part of Fig. 4, we can narrow the spatial distance among the feature vectors associated with different behaviors to make it inseparable, because classifiers rely on this spatial distance to classify [27]. For example, as shown in the right part of Fig. 4, support vector machine (SVM) will construct a hyperplane between different classes at a distance and then classify the samples according to which side of the hyperplane the samples are on. However, more importantly, the ability of the feature vector to be used for identity authentication cannot be excessively affected, *i.e.*, the distance among feature vectors associated with different identities cannot be reduced (see the middle part of Fig. 4). From the perspective of similarity, distance can be characterized by similarity: small distance means high similarity while large distance means low similarity. Thus, it can be derived that eliminating behavior distinguishability is equivalent to improving the similarity among feature vectors associated with different behaviors. Retaining identity distinguishability means not improving the similarity among feature vectors associated with different identities.

We notice that *Siamese network* has outstanding ability in terms of similarity control. Therefore, we design a *Siamese network*-based deep neural network named *BPCloak* to solve the privacy-preserving problem. In the next section, we will detail *BPCloak*'s design, training set construction method, and optimization process.

IV. Privacy Preserving Approach

A. BPCloak Design

The architecture of *BPCloak* is shown in Fig. 5. It is composed of *BPCloak-S* and *BPCloak-Z*. *BPCloak-S* is a feature extraction network, the basic structure of which is *Siamese network*. It is used to extract behavior-irrelevant identity feature vectors. *BPCloak-Z* is an auxiliary identity classifier. It is utilized to improve the identity feature extraction ability of *BPCloak*. Here we introduce the architecture

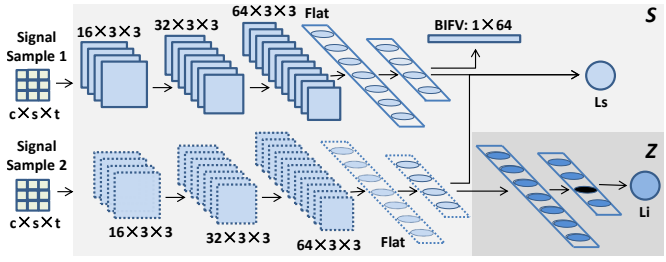


Fig. 5. The architecture of *BPCloak*. It consists of two components: a feature extraction network *BPCloak-S* and an auxiliary identity classifier *BPCloak-Z*.

in detail through the data flow. The inputs of this model are twofold, *i.e.*, pairwise vanilla signal samples $\vec{x_1}$ and $\vec{x_2}$ are fed into *BPCloak* simultaneously. Then $\vec{x_1}$ and $\vec{x_2}$ respectively pass through two five-layer convolutional branches CB_1 and CB_2. These two convolution branches form the whole structure of *BPCloak-S*. They have the same structure and share parameters. Taking branch CB_1 as an example, its first three layers are convolutional layers and each of them consists of three sub-functions, *i.e.*, convolution operation, batch normalization, and rectified linear unit (ReLU) [28]. Among them, the convolution operation is used to extract features from signal samples. The convolution kernel is two-dimensional, and the convolution operation will extract both the spatial feature between sub-channels and the temporal feature between adjacent sampling points. Batch normalization is used to avoid the offset of data distribution and away from the derivative saturation zone. ReLU is used for two purposes: 1) Increasing the nonlinearity of the network to help the network complete the complex filtering task. 2) Making some of the neuron parameters be zero, which makes the network parameters sparse and can reduce the interneuronal dependency. This further optimizes the network's ability to cope with the filtering task. The remaining two layers of CB_1 are fully-connected ones. The output of the third convolutional layer will be flattened into a vector before being fed into the fully connected layer. The first fully-connected layer is followed by a Sigmoid activation function [28] to increase the nonlinearity between fully connected layers. Afterwards, the output of the first fully-connected layer is fed into the second fully-connected layer and becomes $\vec{v_1}$. The dimensionality of $\vec{v_1}$ is $1 \times N_F$, where N_F is the number of neurons in the second fully-connected layer. N_F is set as 64 by default.

While $\vec{x_1}$ has undergone five layers of operations, $\vec{x_2}$ has undergone the same operations and is transformed into $\vec{v_2}$. The $\vec{v_1}$ and $\vec{v_2}$ are considered to be behavior-irrelevant feature vectors (BIFVs). BIFV can be only utilized to identify users and it contains no information about users' behaviors. Henceforth, $\vec{v_1}$ and $\vec{v_2}$ undergo different operations to calculate different losses. In the first operation, $\vec{v_1}$ and $\vec{v_2}$ are simultaneously used to calculate a *similarity loss* (will be introduced in Section IV-C). In the second operation, $\vec{v_2}$ will pass through *BPCloak-Z*, *i.e.*, two fully-connected layers. The output of this layer $\vec{v_p}$ is a probability vector, the elements of which are the probabilities that $\vec{x_2}$ belongs to each user. The $\vec{v_p}$ and the correct identity label of $\vec{x_2}$ are then used to calculate the *identity loss* (will be introduced in Section IV-C). It enables *BPCloak* to extract high-quality identity features from signal samples.

B. Training Set Construction

In order for *BPCloak* to learn the ability to erase behavior information, we need to construct a training set that includes behavior information. As aforementioned, as long as the users appear within the signal's sensing range, their identity information and behavior information will necessarily and concurrently be captured by the signal. The simplest and effective

way to collect signals that meet the requirement is to let users do sensitive behaviors that require protection while they are within the sensing range. In this way, each collected signal sample has two labels: an identity label and a behavior label. Additionally, according to our analysis of 'distinguishability-to-similarity', *BPCloak* needs to change the similarity between two feature vectors, which requires transferring the behavior label and identity label into a similarity label.

In particular, we are interested in two types of grouped pairs (each pair is a training sample): 1) two signal samples with the same identity but different behaviors and 2) two signal samples with the same behavior but different identities. We set the similarity label of the first type of pair to '0' (similar) and set that of the second type of pair to '1' (dissimilar). In this way, *BPCloak* will learn to increase the similarity of the extracted feature vectors with different behaviors and the dissimilarity among extracted feature vectors of different identities will not be reduced. To further boost the extracted identity information contained in the extracted feature vector, we also leverage the original identity labels of signal samples. Finally, each training sample in the training set has two labels: a similarity label and an identity label.

C. Loss Functions and Optimization Process

Given a training set containing n training samples, we form p batches and each batch contains $\frac{n}{p}$ training samples. Each batch is fed into *BPCloak* individually and the losses are calculated based on the outputs. We calculate losses with a unit of batch. For the *similarity loss*, it can be formulated as:

$$\begin{aligned}\mathcal{L}_s =&(1-Y_s)*(D_W(B^S(\vec{x_1}),B^S(\vec{x_2})))^2+\\ &Y_s*(max\left\{0,margin-D_W(B^S(\vec{x_1}),B^S(\vec{x_2}))\right\})^2.\end{aligned} \tag{2}$$

In this formula, $\vec{x_1}$ and $\vec{x_2}$ are two signal samples that belong to the same training sample. Y_s is the similarity label and $margin$ is a distance threshold. $B^S(\cdot)$ represents *BPCloak-S* and $D_W(B^S(\vec{x_1}),B^S(\vec{x_2}))$ is the *Euclidean distance* between the BIFVs $\vec{v_1}$ and $\vec{v_2}$, where $\vec{v_1}=B^S(\vec{x_1})$ and $\vec{v_2}=B^S(\vec{x_2})$. The $margin$ is set as 3 empirically. If we denote $\vec{v_1}$ and $\vec{v_2}$ as $[v_1^1,v_1^1,\cdots,v_1^k]$ and $[v_2^1,v_2^1,\cdots,v_2^k]$, their *Euclidean distance* can be calculated by:

$$\sqrt{(v_1^1-v_2^1)^2+(v_1^2-v_2^2)^2+\cdots+(v_1^k-v_2^k)^2}. \tag{3}$$

As for the *identity loss*, we opt to use *cross-entropy loss* [29]. It can be formulated as:

$$\mathcal{L}_i=-\sum_{c=1}^{M}y_c\log(P_c), \tag{4}$$

in which y_c is the indication variable, P_c is the probability that the signal sample belongs to identity label c and M is the maximum of the identity labels. Ultimately, the final loss consisting of the *similarity loss* and *identity loss* can be formulated as:

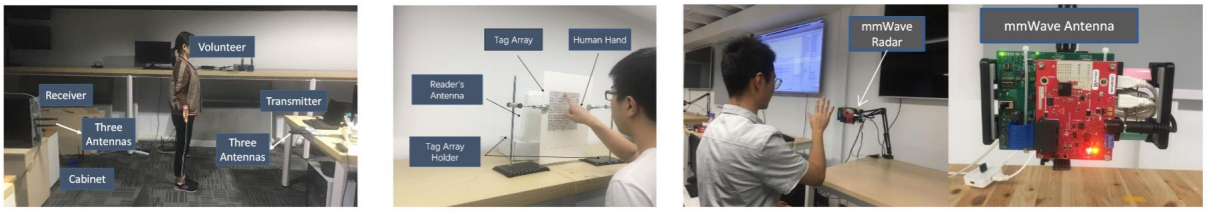
$$\mathcal{L}_f=\alpha*\mathcal{L}_s+(1-\alpha)*\mathcal{L}_i,\quad \alpha\in(0,1). \tag{5}$$

By updating the parameters in *BPCloak* using $\mathcal{L}_f$, *BPCloak* can learn to reduce the distinguishability between two signal samples with different behaviors, while remaining the distinguishability between two signal samples with the same identity. Thus, a trained *BPCloak* is able to remove behavior information from vanilla signal samples while retaining sufficient identity information.

V. Evaluation and Result

In this section, we conducted experiments on three real-world systems (*i.e.*, a COTS WiFi system, a COTS RFID system, and a COTS mmWave system). The condition for the adversary to achieve the best-case attack is that the adversary can use a part of the user's signal samples and corresponding behavior labels to train the malicious classifier. Therefore, to effectively evaluate our privacy-preserving method, all experiments were carried out under this condition.

Experiment setup: 1) WiFi system: we invited ten volunteers (two females and eight males) to perform ten gestures in the WiFi sensing range. The ages of volunteers varied from 22 to 35 and their heights varied from 160cm to 188cm. The experiment setup is shown in Fig. 6(a), the transmitter equipped with an *Atheros 9380* network interface card (NIC) was placed 2m away from the receiver (also equipped with an *Atheros 9380* NIC). Both the transmitter and the receiver were placed on wooden furniture, the top surfaces of which were 80cm off the ground. Ten gestures representing ten numbers from zero to nine are shown in Fig. 7. When posing gestures, the volunteer was standing between the transmitter and the receiver. We totally collected over 29000 signal samples in the WiFi system. 2) RFID system: we invited five volunteers (two females and three males) to perform ten activities. The ages of the volunteers varied from 21 to 31 and the heights of them varied from 165cm to 188cm. As shown in Fig. 6(b),



(a) WiFi. (b) RFID. (c) Millimeter wave.

Fig. 6. The experiment setups of three systems.

TABLE I
THE AUTHENTICATION ACCURACY AND RECOGNITION ACCURACY OF WIFI DOP AND BIFV AMONG DIFFERENT CLASSIFIERS.

Methods	RF	LR	KNN	SVM	DT	NB	NN
AA-DoP	99.84%	99.96%	99.63%	99.56%	97.83%	98.39%	99.79%
RA-DoP	15.21%	19.09%	73.86%	83.80%	75.54%	73.93%	93.69%
AA-BIFV	99.45%	99.40%	99.35%	99.37%	99.24%	99.34%	97.69%
RA-BIFV	10.49%	10.87%	10.51%	10.66%	10.42%	10.95%	10.02%

TABLE II
THE AUTHENTICATION ACCURACY AND BEHAVIOR RECOGNITION ACCURACY OF RFID DOP AND BIFV AMONG DIFFERENT CLASSIFIERS.

Methods	RF	LR	KNN	SVM	DT	NB	NN
AA-DoP	99.60%	95.44%	99.50%	93.35%	62.60%	63.89%	100.00%
RA-DoP	95.36%	53.10%	95.63%	94.64%	48.12%	27.38%	89.78%
AA-BIFV	97.63%	97.71%	97.68%	97.74%	97.00%	95.81%	76.20%
RA-BIFV	17.95%	12.45%	15.37%	12.27%	17.44%	11.46%	10.75%

TABLE III
THE AUTHENTICATION ACCURACY AND RECOGNITION ACCURACY OF MMWAVE DOP AND BIFV AMONG DIFFERENT CLASSIFIERS.

Methods	RF	LR	KNN	SVM	DT	NB	NN
AA-DoP	77.36%	99.57%	59.33%	99.97%	36.27%	78.24%	99.70%
RA-DoP	38.92%	86.41%	32.04%	93.15%	18.81%	38.68%	93.68%
AA-BIFV	96.05%	94.94%	94.57%	95.18%	93.62%	84.79%	94.71%
RA-BIFV	27.22%	9.62%	22.24%	19.47%	21.42%	10.52%	10.19%

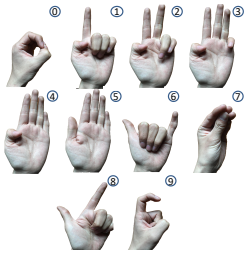


Fig. 7. The gestures of zero to nine.

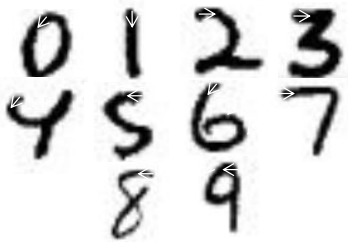


Fig. 8. The activities of zero to nine.

the tag array was formed by 49 tags (*Alien-9629*), and the reader's antenna (*Impinj R420+Larid A9028*) is placed 10cm away from the tag array. The activities of zero-to-nine are shown in Fig. 8. The position of the white arrow is the starting position of writing and its direction is the starting direction of writing. Volunteers were asked to write ten numbers in front of the tag array. We totally collected over 4000 signal samples in the RFID system. 3) Millimeter wave system: we invited nine volunteers (seven males and two females) to pose ten gestures shown in Fig. 7. The ages of the volunteers varied from 22 to 24 and the heights of them varied from 160cm to 175cm. As shown in Fig. 6(c), We asked volunteers to sit in front of the mmWave radar (*IWR1642*) to pose gestures. The radar was connected to one transmission antenna and four receiving antennas. We totally collected 9000 signal samples in the mmWave system.

Data pre-processing: The received WiFi signals were CSI data, the elements of which were complex numbers. We first utilized a low-pass filter (with a cutoff frequency of 20Hz) to remove the high-frequency noise in the CSI data. Then we calculated the absolute value of the CSI data to obtain the amplitude of the CSI. Afterwards, we segmented the time-series data of each sub-channel for each gesture so that each signal sample (which is associated with a specific gesture and a specific volunteer) had a dimensionality of $1 \times 504 \times 10$. In the RFID system, the received signals were also time-series data for each tag. We first selected the first 30 sampling points for each tag and then concatenated them together. In this way, each signal sample had a dimensionality of $2 \times 49 \times 30$. In the mmWave system, we also calculated the absolute values (*i.e.*, amplitudes) of collected signals. The dimensionality of each signal sample is $1 \times 4 \times 1024$.

Metrics: We defined three metrics to quantify the attack effectiveness of the behavior privacy mining and the privacy-preserving performance of *BPCloak*: accuracy, defense rate (DR), and trade-off rate (TR). The accuracy is the probability that the identity/behavior label of any signal sample is correctly identified. It is termed as authentication accuracy (AA) in identity authentication, and recognition accuracy (RA) in behavior recognition. It can be formulated as: $accuracy = \frac{N_{cor}}{N_{all}}$, where N_{cor} is the number of correctly classified test signal samples and N_{all} is the number of all test samples. DR is the ratio of the RA that *BPCloak* reduces to the RA that the adversary can achieve. The larger the DR, the better the privacy-preserving performance of *BPCloak*. DR can be formulated as:

$$DR = \frac{RA_{att} - RA_{def}}{RA_{att}}, \tag{6}$$

where RA_{att} is the RA of vanilla signal samples and RA_{def} is the RA of extracted BIFVs. TR is the ratio of the AA that *BPCloak* loses during the privacy preserving to the AA of the vanilla signal samples. The smaller the TR, the better the

performance of *BPCloak*. TR can be formulated as:

$$TR = \frac{AA_{ori} - AA_{def}}{AA_{ori}}. \quad (7)$$

In this formula, AA_{ori} is the AA of vanilla signal samples and AA_{def} is the AA of extracted BIFVs.

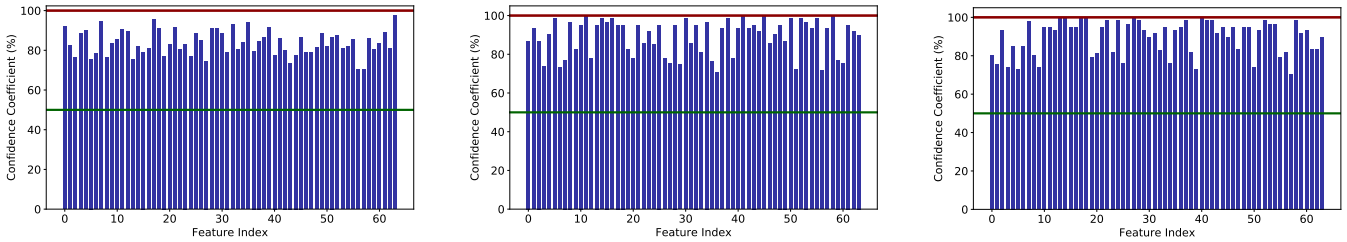
A. Overall Performance

For the sake of training *BPCloak*, we first constructed training set with one thousand randomly-selected training samples for each system. Then, in order to ensure that the experiment results will not be affected by the randomness of the training set construction method, we train the *BPCloak* by using these training sets and their subsets in the following experiments. At the same time, to ensure that the signal samples involved in the *BPCloak* training and the signal samples used to extract BIFVs do not overlap, the data we used to extract BIFV consists of other signal samples that are not in the training set. Hereinafter, the vanilla signal samples used to extract BIFVs are called the data to be protected (DoP). The RA_{att} and AA_{ori} in Eq. 6 and Eq. 7 are the RA and AA of DoP.

RA and AA of DoP: We first calculated the RA and AA of DoP. It should be noted that in the WiFi system, RA is closely related to the person [2], *i.e.*, the mixed signal samples of multiple people cannot be used for multi-person behavior recognition with high accuracy. So, in the WiFi experiment, we calculated the RA of each person individually and then calculated the mean of these RA. We used mainstream machine learning classifiers that existing identity authentication and behavior recognition works used: random forest (RF), LR, k-nearest neighbors (KNN), SVM, decision tree (DT), naive Bayes (NB), and three-layer neural network (NN). In all experiments, we randomly selected 75% of the DoP/BIFVs to train the classifier and used the remaining 25% of the DoP/BIFVs to calculate the accuracy. This process was repeated ten times and we took the average of ten accuracy as the final result. The experimental results of the WiFi system, the RFID system, and the mmWave system are shown in the top two rows in Table I, II, and III respectively. The results of the WiFi system indicate that the best attack effectiveness (the highest RA) exceeds 99%. Meanwhile, an SP can use the DoP to provide an AA of 99%+. In both RFID and mmWave systems, an SP can use the DoP to provide high AA (99%+). The adversary's possible attack effectiveness of the RFID system and mmWave system exceed 94% and 99% respectively. These experiment results show that although SP can use the DoP to provide users with accurate identity authentication, an adversary may use the DoP to monitor their behaviors.

RA and AA of BIFV: Then, we evaluated the RA and AA of the BIFVs extracted from the DoP. The results of the WiFi, the RFID, and the mmWave systems are shown in the bottom two rows in Table I, II, and III, respectively. It can be observed that the AA of BIFVs are not much lower than that of the DoP, or even increase on some classifiers. But RA is greatly reduced on all classifiers. The RA in the WiFi system even approximates random guess (10%). Therefore, the identity information is well retained in the BIFV, so an SP can still provide user authentication with high accuracy. At the same time, the adversary cannot achieve decent attack effectiveness even under the best attack condition.

Chi-Square test: In order to verify from a statistical level that BIFVs are irrelevant to the behavior information,



(a) WiFi. (b) RFID. (c) Millimeter wave.

Fig. 9. The confidence coefficients for accepting the hypothesis that BIFV is irrelevant to the behavior information. The green horizontal line means 50% and the red horizontal line means 100.00%.

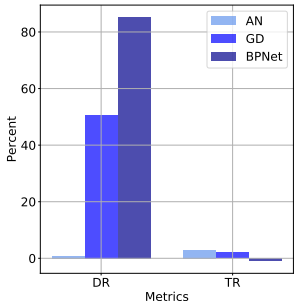


Fig. 10. The DR and TR of the WiFi system.

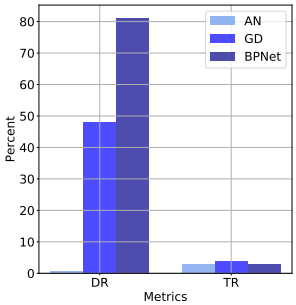


Fig. 11. The DR and TR of the RFID system.

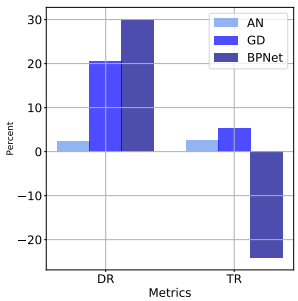


Fig. 12. The DR and TR of the mmWave system.

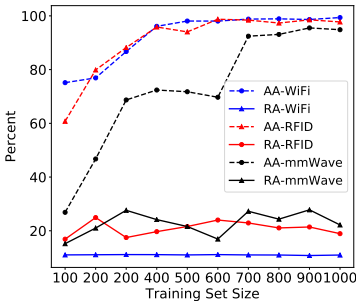


Fig. 13. The performance among different training set sizes.

we performed a Chi-Square independence test [30] towards each element position in BIFV (64 positions by default). Specifically, we removed every position of the element in BIFV alternatively and examined the effect of BIFV on RA after removing the element at that position. Our hypothesis to be validated is: the element at this position is independent of RA, *i.e.*, the element at this position is irrelevant to the behavior information. In each system, we randomly selected 500 BIFVs and counted the number of BIFVs that are correctly and incorrectly classified towards behavior before and after removing an element. With a degree of freedom of one, we show the confidence coefficients of the three systems to accept the hypothesis in Fig. 9(a), (b), and (c), respectively. The experiment results show that all the confidence coefficients are larger than 50% or even 70%. Some of the confidence coefficients even approximate 100%. This result demonstrates that we have a high degree of confidence to accept the hypothesis that the BIFV is irrelevant to the behavior information. Therefore, *BPCloak* has efficient behavior privacy-preserving capabilities and it is extremely difficult for an adversary to recover the behavior privacy.

DR and TR: According to the experiment results in Table I, II, and III, NB has the highest RA on WiFi BIFVs, and RF has the highest RA on RFID and mmWave BIFVs. Therefore, we used NB and RF to evaluate the DR and TR of *BPCloak*. We compared *BPCloak* with two other deep learning-based baselines. One is the adversarial network (AN) [9]. Its principle is to use a generator to extract BIFV, the other two discriminators are used to reduce the loss of identity authentication and increase the loss of behavior recognition. The other one is gradient descent (GD) [31]. We first randomly generate a perturbation, and then update the perturbation based on reducing the loss of identity recognition and increasing the loss of behavior recognition. The comparison results of the WiFi system, the RFID system, and the mmWave system are shown in Fig. 10, 11, and 12, respectively. The experiment results indicate that whether it is in the WiFi system, RFID system, or mmWave system, the DR of *BPCloak* is higher than that of other two baselines. This demonstrates that *BPCloak* can effectively erase the behavior information in DoP and it outperforms the other two baselines. At the same time, in the three systems, the TRs of *BPCloak* are lower than that of other two baselines. In the WiFi and mmWave systems, the TRs of *BPCloak* are even negative, which means that *BPCloak* performs better than the other two baselines in identity information retention. More importantly, *BPCloak* can effectively retain the identity information in DoP and even make it more prominent.

B. The Effect of Training Set Size

In order to explore the effect of the training set size (*i.e.*, the number of training samples in the training set) on the performance of *BPCloak*, we varied the training set size from 100 to 1000 with a stride of 100. The AA and RA of the BIFVs of the three systems are shown in Fig. 13. It can be seen from the experiment results that the variation trends of the AA and RA are basically the same in the three systems. RA basically remains stable. AA increases as the training set size increases. When the size of the training set is 700, the AA curves of the three systems become flat. Henceforth, the AA of the BIFVs of the three systems are all significantly high and the RA is low. Therefore, only a small training set is needed to train an effective *BPCloak*.

C. Transferability Evaluation

In this experiment, we explored the transferability of *BPCloak* in terms of unseen identities and behaviors. In specific, we first used a trained *BPCloak* to extract the BIFVs of the identities (unseen identities) not participating in the training and then classified these BIFVs. Then, we used the trained *BPCloak* to extract the BIFVs of the behaviors (unseen behaviors) not participating in the training and recognized them. In the identity experiment, we used two identities as unseen identities and used the signal samples of other identities to train *BPCloak*. The AA of unseen identities are 99.13%, 82.14%, and 100.00% in the WiFi system, RFID system, and mmWave system, respectively. The AA of the RFID system is relatively low because the training set used to train *BPCloak* has only three persons' data, which makes *BPCloak* not learn the expected capability for identity information retention. Nevertheless, the results still prove that *BPCloak* has outstanding transferability for new users. In the behavior experiment, we used the signal samples of the first five behaviors to train *BPCloak* and then extract the BIFVs of the signal samples of the other five behaviors. The experiment results show that the RA of unseen behaviors are 23.60%, 34.58%, and 69.99% in the WiFi, the RFID, and the mmwave systems, respectively. The results show that *BPCloak* still has a strong ability to erase the privacy of unseen behaviors in the WiFi system and RFID system. In the mmWave system, *BPCloak* has relatively weaker privacy protection capabilities compared

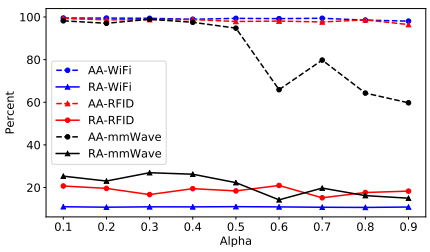


Fig. 14. The effect of the α.

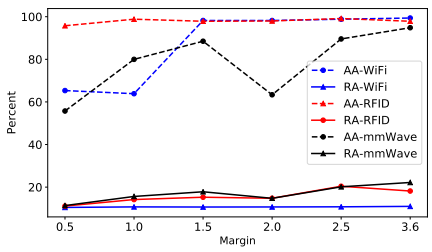


Fig. 15. The effect of the *margin*.

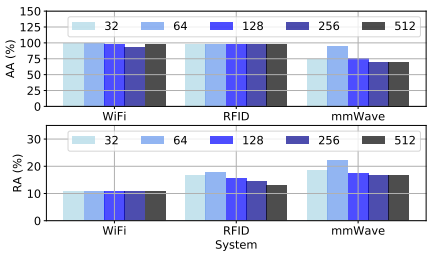


Fig. 16. The effect of the BIFV size.

with the other two systems. This result shows that *BPCloak* also has outstanding transferability for unseen behaviors.

D. Ablation Study

In this part, we explored the effect of hyperparameters α, *margin*, and feature vector size (the number of elements in BIFV) on the privacy-preserving capability of *BPCloak*.

The effect of α: In this experiment, we varied the value of α from 0.1 to 0.9 in steps of 0.1. The experiment results in Fig. 14 show that the variation of α has no obvious effect on the RA of BIFV. However, a smaller α can better retain identity information. In the mmWave system, a large α will make the AA of the BIFV unstable. So 0.2 is a better choice for α.

The effect of *margin*: In this experiment, we varied the value of *margin* from 0.5 to 3.0 in steps of 0.5. The experiment results in Fig. 15 show that the increase of $margin$ will increase both the AA and the RA of BIFV. However, the $margin$ has a much larger effect on AA than RA. In order to achieve a significantly high AA, the $margin$ is best to take 3.0.

The effect of BIFV size: In this experiment, we selected five feature sizes in common used: 32, 64, 128, 256, and 512. The AA and RA are shown in the upper part and the lower part of Fig. 16, respectively. It can be found that in the WiFi and RFID systems, the variation in feature size does not have an obvious effect on AA. In the mmWave system, as the feature vector size increases, the AA of BIFV first increases and then decreases. The maximum value is reached when the feature size is 64. As for the RA, it does not change markedly in the WiFi system. But it increases first and then decreases in both RFID and mmWave systems. It also reaches the maximum when the feature size is 64. However, even the maximum value is only close to 20%. Therefore, if in order to provide high-quality identity authentication service as much as possible while protecting behavior privacy acceptably, 64 is the best choice.

VI. Related Work

In this section, we first introduce the RF signal-based identity authentication technology. Then we introduce some privacy-preserving techniques.

RF signal-based identity authentication: Existing authentication technologies based on RF signals generally follow the same authentication mode, that is, first extract feature vectors, and then use learning-based classifiers to distinguish feature vectors. For example, RF-Rhythm [32] extracts the user's tapping rules on the tag array as a feature vector, and uses NN, SVM, and CNN three classifiers for identity recognition. WiPIN [4] extracts the human body features in the WiFi signal as a feature vector after the signal passes through the human body, and finally uses SVM for identification. MU-ID [33] leverages mmWave radar to capture the limb motions and gaits of multiple users as features. Then it uses CNN to achieve multi-user identification. Different from previous works, *BPCloak* not only pursues high-quality identity authentication but also ensures that behavior privacy is not leaked.

Privacy-preserving technique: Such techniques usually leverage perturbation or machine learning to achieve the privacy-preserving goal. For instance, in [34], Rezaei *et al.* utilize a generative adversarial net to generate perturbation. The perturbation is then added to the privacy-sufficient data to protect users' private individual attributes. To protect database privacy, Dwork *et al.* [35] add certain noise to the data. Bayerl *et al.* [36] realize privacy-preserving automatic speaker verification by outsourced secure two-party computation. To prevent mobile devices from re-identification attacks while retaining the utility of activity, Jourdan *et al.* [37] propose a machine learning-based framework to reduce the re-identification accuracy. However, there has been very little work to protect the privacy of RF signals. We thus proposed a promising deep learning-based method to erase the behavior privacy contained in RF signals.

VII. Discussion and Future Work

In the evaluation section, we apply *BPCloak* to identity authentication. However, there are many applications based on RF signals, such as user tracking and localization. It has been proved that the same RF signal can be used to locate and recognize behavior [38]. Fortunately, *BPCloak* can be easily extended to other applications. For example, in RF-based localization, we only need to select signals belonging to the same behavior in different locations and signals with different behaviors at the same location to construct a training set at first. Then we transform *BPCloak* into the architecture shown in Fig 17. The difference between the authentication-oriented *BPCloak* and the localization-oriented *BPCloak* is that their second losses are different. In the localization-oriented *BPCloak*, the second loss is the *localization loss* ($\mathcal{L}_l$). It calculates the absolute value of the difference between the two output coordinates (X, Y) and the two ground-truth coordinates (X_g, Y_g). In this way, *BPCloak* can protect user behavior privacy in localization applications. Thus, *BPCloak* has decent scalability in theory. We will evaluate such scalability in future work.

VIII. Conclusion

In this paper, we show the risk of behavior privacy leakage in RF-based user authentications. To address it, we designed a novel network named *BPCloak*. *BPCloak* can effectively filter behavior information out from RF signals without impairing the ability of signals to be used for accurate identity authentication. We conducted comprehensive experiments on

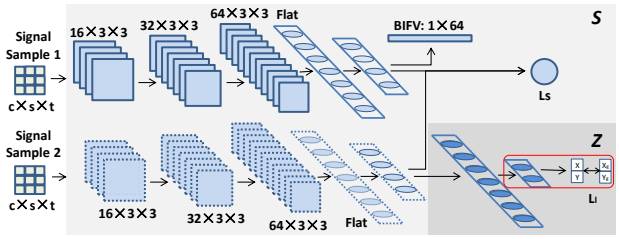


Fig. 17. The architecture of localization-oriented *BPCloak*.

three real-world systems. The experimental results show that *BPCloak* can effectively hide behavior information.

IX. Acknowledgement

This work is supported in part by National Key Research and Development Program of China (Grant No. 2020AAA0107700), National Natural Science Foundation of China (Grants No. 61872285, 62032021, 61772236, and 61872081). Zhejiang Key Research and Development Plan (Grant No. 2019C03133), Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (Grant No. 2018R01005), Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Research Institute of Cyberspace Governance in Zhejiang University.

This work is supported in part by National Key Research and Development Program of China (Grant No. 2020AAA0107700), National Natural Science Foundation of China (Grants No. 61872285, 62032021, 61772236). Zhejiang Key Research and Development Plan (Grant No. 2019C03133), Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (Grant No. 2018R01005)

References

[1] C. Zhao, Z. Li, T. Liu, H. Ding, J. Han, W. Xi, and R. Gui, "Rf-mehndi: A fingertip profiled RF identifier," in *IEEE Conference on Computer Communications, INFOCOM*, 2019.

[2] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *International Conference on Mobile Systems, Applications, and Services, MobiSys*, 2019.

[3] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single wi-fi link," in *International Conference on Mobile Systems, Applications, and Services, MobiSys*, J. Ott, F. Dressler, S. Saroiu, and P. Dutta, Eds., 2018.

[4] F. Wang, J. Han, F. Lin, and K. Ren, "Wipin: Operation-free person identification using wifi signals," in *IEEE Global Communications Conference, GLOBECOM*, 2019.

[5] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "Freesense: Indoor human identification with wi-fi signals," in *IEEE Global Communications Conference, GLOBECOM*, 2016.

[6] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN*, 2016.

[7] J. Liu, X. Zou, J. Han, F. Lin, and K. Ren, "BioDraw: Reliable multi-factor user authentication with one single finger swipe," in *IEEE/ACM International Symposium on Quality of Service, IWQoS*, 2020.

[8] X. Ma, J. Qu, J. Li, J. C. S. Lui, Z. Li, and X. Guan, "Pinpointing hidden iot devices via spatial-temporal traffic fingerprinting," in *IEEE Conference on Computer Communications, INFOCOM*, 2020.

[9] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *International Conference on Mobile Computing and Networking, MobiCom*, 2018.

[10] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT*, vol. 2, no. 1, pp. 23:1–23:21, 2018.

[11] P. O. POSITIVITY, "What do your daily habits reveal about your personality?" https://www.powerofpositivity.com/personality-habits/, 2020.

[12] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *International Conference on Mobile Computing and Networking MobiCom*, 2016.

[13] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "When csi meets public wifi: Inferring your mobile phone password via wifi signals," in *ACM SIGSAC Conference on Computer and Communications Security, CCS*, 2016.

[14] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, 2005.

[15] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016.

[16] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing, TMC*, vol. 16, no. 2, pp. 581–594, 2017.

[17] D. Vasisht, G. Zhang, O. Abari, H. Lu, J. Flanz, and D. Katabi, "In-body backscatter communication and localization," in *Conference of the ACM Special Interest Special Interest Group on Data Communication, SIGCOMM*, 2018.

[18] Y. Song, Z. Cai, and Z. Zhang, "Multi-touch authentication using hand geometry and behavioral information," in *IEEE Symposium on Security and Privacy, S&P*, 2017.

[19] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, "Learning longterm representations for person re-identification using radio signals," in *IEEE Conference on Computer Vision and Pattern Recognition,CVPR*, 2020.

[20] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *International Conference on Distributed Computing in Sensor Systems, DCOSS*, 2016.

[21] Y. Wang and Y. Zheng, "Modeling RFID signal reflection for contact-free activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT*, vol. 2, no. 4, pp. 193:1–193:22, 2018.

[22] C. Li, M. Liu, and Z. Cao, "Wihf: Enable user identified gesture recognition with wifi," in *IEEE Conference on Computer Communications, INFOCOM*, 2020.

[23] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[24] R. M. C. R. de Souza, D. C. F. Queiroz, and F. J. de A. Cysneiros, "Logistic regression-based pattern classifiers for symbolic interval data," *PATTERN ANALYSIS AND APPLICATIONS*, vol. 14, no. 3, pp. 273–282, 2011.

[25] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "802.11 with multiple antennas for dummies," *Computer Communication Review*, vol. 40, no. 1, pp. 19–25, 2010.

[26] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-wifi: Fine-grained person perception using wifi," in *IEEE/CVF International Conference on Computer Vision, ICCV*, 2019.

[27] B. Z. H. Zhao, H. J. Asghar, and M. A. Kaafar, "On the resilience of biometric authentication systems against random inputs," in *Network and Distributed System Security Symposium, NDSS*, 2020.

[28] A. M. Pretorius, E. Barnard, and M. H. Davel, "Relu and sigmoidal activation functions," in *South African Forum for Artificial Intelligence Research*, 2019.

[29] K. Nar, O. Ocal, S. S. Sastry, and K. Ramchandran, "Cross-entropy loss and low-rank features have responsibility for adversarial examples," *CoRR*, vol. abs/1901.08360, 2019.

[30] STATOLOGY, "Chi-square test of independence: Definition, formula, and example." https://www.statology.org/chi-square-test-of-independence/, 2020.

[31] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *International Conference on Learning Representations, ICLR*, 2018.

[32] J. Li, C. Wang, A. Li, D. Han, Y. Zhang, J. Zuo, R. Zhang, L. Xie, and Y. Zhang, "Rf-rhythm: Secure and usable two-factor RFID authentication," in *IEEE Conference on Computer Communications, INFOCOM*, 2020.

[33] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "MU-ID: multi-user identification through gaits using millimeter wave radios," in *IEEE Conference on Computer Communications, INFOCOM*, 2020.

[34] A. Rezaei, C. Xiao, J. Gao, and B. Li, "Protecting sensitive attributes via generative adversarial networks," *CoRR*, vol. abs/1812.10193, 2018.

[35] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *International Conference on the Theory and Applications of Cryptographic Techniques*, 2006.

[36] Sebastian P Bayerl, Ferdinand Brasser, Christoph Busch, Tommaso Frassetto, Patrick Jauernig, Jascha Kolberg, Andreas Nautsch, Korbinian

Riedhammer, Ahmad-Reza Sadeghi, Thomas Schneider, Emmanuel Stapf, Amos Treiber, and Christian Weinert, "Privacy-preserving speech processing via STPC and TEEs," 11 2019.

[37] T. Jourdan, A. Boutet, and C. Frindel, "Toward privacy in iot mobile devices for activity recognition," in *EAI International Conference on Mobile and Ubiquitous Systems, EAI MobiQuitous*, 2018.

[38] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with wifi fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.