Generative AI Based Secure Wireless Sensing for ISAC Networks

Jiacheng Wang, Hongyang Du, *Member, IEEE*, Yinqiu Liu, Geng Sun, *Senior Member, IEEE*, Dusit Niyato, *Fellow, IEEE*, Shiwen Mao, *Fellow, IEEE*, Dong In Kim, *Life Fellow, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Integrated sensing and communications (ISAC) is one of the crucial technologies for 6G, and channel state information (CSI) based sensing serves as an essential part of ISAC. However, current research on ISAC focuses mainly on improving sensing performance, overlooking security issues, particularly the unauthorized sensing of users. Hence, this paper proposes a diffusion model based secure sensing system (DFSS). Specifically, we first propose a discrete conditional diffusion model to generate graphs with nodes and edges, which guides the ISAC system to appropriately activate wireless links and nodes, ensuring the sensing performance while minimizing the operation cost. Using the activated links and nodes, DFSS then employs the continuous conditional diffusion model to generate safeguarding signals, which are next modulated onto the pilot at the transmitter to mask fluctuations caused by user activities. As such, only authorized ISAC devices with the safeguarding signals can extract the true CSI for sensing, while unauthorized devices are unable to perform the effective sensing. Experiment results demonstrate that DFSS can reduce the activity recognition accuracy of the unauthorized devices by approximately 70%, effectively shield the user from the illegitimate surveillance.**

***Index Terms*—Generative AI, integrated sensing and communication, wireless sensing security.**

I. Introduction

INTEGRATED Sensing and Communication (ISAC) is an emerging technology, which integrates wireless sensing and communication into a single system, effectively utilizing the network resources to simultaneously perform data exchange and environmental sensing [1]. Such integration not only enhances the spectral and energy efficiency, but also reduces operation and hardware costs, offering broad applications in various scenarios [2], [3], [4], [5]. A representative example of ISAC is the channel state information (CSI) based sensing [6], which has seen rapid advancement over the last decade. This technology involves the processing and analysis of CSI from wireless communications networks to sense humans, including their locations, activities, and even breathing and heart rates [7].

In CSI-based sensing, the research is comprehensive but faces major security weaknesses. Rigorously, these systems rely on the CSI, which is derived based on the communication protocols and pilot signals shared between the signal transmitter and receiver in wireless communications networks [8]. This implies that any standard wireless device in open spaces can intercept these signals and use the standardized pilot signals to measure the CSI. By processing and analyzing CSI measurements, unauthorized devices can then glean insights into an individual's daily activities [9]. For example, an unauthorized access point (AP) could capture Wi-Fi signals and extract CSI to deduce the gait information, posing a serious threat to individual privacy [10]. Another major concern is the vulnerability to attacks from rogue APs, such as using customized APs to capture and modify signal amplitude and phase to create spoofing signals. These spoofing signals are transmitted to disrupt legitimate sensing, adversely impacting the performance of the ISAC system [11]. Figure 1 depicts CSI-based sensing and two primary security challenges in CSI sensing systems, emphasizing the crucial need to protect individuals' daily activities from unauthorized AP monitoring.

To shield users from unauthorized surveillance, an effective approach is to design and generate a safeguarding signal and modulate it onto the pilot signals for CSI estimation, thereby masking the signal fluctuations caused by the user activity. Here, it is crucial that this safeguarding signal remains

Received 19 August 2024; revised 31 March 2025; accepted 29 April 2025. Date of publication 14 May 2025; date of current version 3 June 2025. This work was supported in part by the National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research and Development Program, under Grant FCP-NTU-RG-2022-010 and Grant FCP-ASTAR-TG-2022-003; in part by Singapore Ministry of Education (MOE) Tier 1 under Grant RG87/22 and Grant RG24/24; in part by the Nanyang Technological University (NTU) Centre for Computational Technologies in Finance (NTU-CCTF) and the RIE2025 Industry Alignment Fund–Industry Collaboration Projects (IAF-ICP), administered by Agency for Science, Technology and Research (A*STAR) under Award I2301E0026; in part by the Alibaba Group and NTU Singapore through Alibaba-NTU Global e-Sustainability CorpLab (ANGEL); and in part by the National Research Foundation of Korea (NRF) Grant funded by Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2021R1A2C2007638. The work of Geng Sun was supported in part by the National Natural Science Foundation of China under Grant 62272194 and Grant 62471200. The work of Shiwen Mao was supported in part by NSF under Grant CCSS-2434053 and Grant CNS-2107190. The associate editor coordinating the review of this article and approving it for publication was Prof. Linke Guo. *(Corresponding authors: Geng Sun; Hongyang Du.)*

Jiacheng Wang, Yinqiu Liu, and Dusit Niyato are with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798 (e-mail: jiacheng.wang@ntu.edu.sg; yinqiu001@ntu.edu.sg; dniyato@ntu.edu.sg).

Hongyang Du is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: duhy@eee.hku.hk).

Geng Sun is with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: sungeng@jlu.edu.cn).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Dong In Kim is with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea (e-mail: dongin@skku.edu).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TIFS.2025.3570202

1556-6021 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

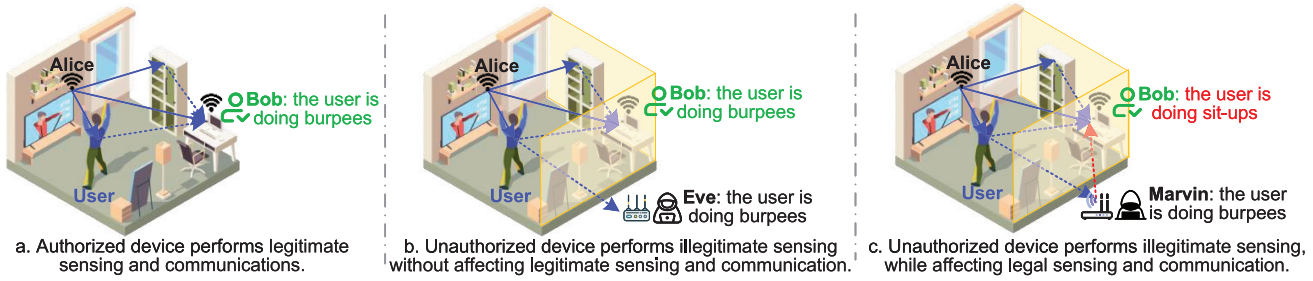


Fig. 1. The principles of CSI based sensing and two main types of security issues. Here, Alice is the ISAC device which transmits the wireless signal, and Bob is the authorized device, which performs legitimate sensing of the user. Eve is an unauthorized eavesdropper who performs illegitimate sensing of the user by capturing and processing the signals. Marvin is an unauthorized attacker who can not only sense the users, but also generate spoofing signals to disrupt legitimate communications and sensing.

unreconstructable by unauthorized devices. Moreover, the signal generation must adapt to changes in external conditions, thereby ensuring universality. The recent progress in artificial intelligence generated content (AIGC) [12] has led to the development of generative AI (GAI) models, providing support for the safeguarding signal generation [13].

Within the field of GAI, diffusion models have gained particular attention as they can generate high-quality text, images, and videos based on user prompts [14], providing solid support for applications such as DALL-E. Furthermore, diffusion models also excel in signal denoising [15] and generation [16], delivering many unique advantages. First, they demonstrate creativity, capable of producing data that resembles yet differs from the training data, thereby ensuring diversity. Second, they generate data through a denoising, which relies on specific network parameters and random seeds, leading to unpredictable results. Furthermore, they can fine-tune the generation process to match user prompts or generation condition closely [17], exhibiting considerable adaptability. These characteristics make diffusion models a suitable choice for safeguarding signal generation.

Building on the discussion above, we propose a diffusion model based secure sensing system (DFSS). Specifically, we first propose a discrete conditional diffusion model to generate the graph with nodes and edges. This graph guides the ISAC system to activate wireless links appropriately for sensing, ensuring performance while minimizing link activation for cost-effective operation. Then, DFSS employs a continuous conditional model to generate safeguarding signals. Compared to methods based on predefined algorithms and codebook design, the generated safeguarding signals show greater randomness and diversity, making them harder to replicate. These signals are then modulated onto the pilot symbols at the transmitter to mask fluctuations caused by user activities. At the receiver, authorized devices use the same model and seed to generate identical safeguarding signals, enabling them to extract real CSI for communication and sensing. However, unauthorized devices are unable to generate the safeguarding signal, and hence fail to extract CSI for sensing. In summary, the contributions of this paper are as follows.

- We propose a discrete conditional diffusion model (D-CDM) to generate graphs by using ISAC device distribution and user locations as conditions. These graphs guide appropriate activation of wireless links and nodes, ensuring optimal sensing performance while minimizing the costs for efficient ISAC network operation.
- We design safeguarding signals based on the characteristics of signal fluctuations caused by human indoor activities. We develop a dataset to train a continuous diffusion model. This enables the generation of safeguarding signals based on activated links and nodes and the user to be protected. The generated signals exhibit diversity and randomness, making them challenging for unauthorized devices to replicate.
- We propose to modulate the safeguarding signals onto the pilot at the transmitter, which prevents unauthorized devices from extracting the real CSI for sensing. Compared to existing methods, our approach does not require additional equipment or prior information such as the location of unauthorized devices, making it more cost-effective and practical.
- We evaluate the DFSS using software-defined radio devices. Taking activity recognition as an example, the results demonstrate that DFSS can reduce the activity recognition accuracy of the state-of-the-art system by about 70%, confirming its effectiveness in protecting users from illegitimate sensing.

II. Related Work

In this section, we review existing CSI based sensing systems and analyze the security issues within them.

A. CSI Based Sensing

With the proliferation of wireless devices, CSI-based sensing technologies have been significantly advanced. Researchers have proposed various CSI-based sensing systems, including localization [18], [19], activity and gesture recognition [8], [20], respiratory and heartbeat monitoring [21], [22], and authentication [23]. In [24], the authors trained the weights of the deep network as fingerprints and employed a greedy learning to reduce complexity. Then, they used a radial basis function-based probabilistic approach to achieve localization, with a mean error of approximately 0.94 m. The authors in [13] employed clustering to analyze the multi-domain parameters of reflections caused by humans. This facilitated the human flow detection, especially the number of subflows and subflow sizes, achieving the detection accuracy of 92% and 91%, respectively. For activity recognition, the

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

authors in [25] proposed an attention-based bi-directional long short-term memory network to learn features from sequential CSI measurements in two directions, achieving an overall recognition accuracy of 97.3%. For small scale sensing, the authors in [22] derived breathing signals from the CSI phase difference between antennas, and then integrated the signal matching algorithm with the peak detection to estimate respiratory rates. In single-user scenarios, evaluations show an estimation error of less than 0.5 bpm in 96% of cases.

B. Security in CSI Based Sensing

In a CSI sensing system, there are two main types of security issues. **The first type** involves unauthorized devices performing illegitimate sensing to obtain user's relevant information [26], as shown in part *b* of the Fig. 1. For instance, the authors in [27] proposed a signal propagation model that converts eavesdropped CSI into those of legitimate devices. Then, they added user identification models into the substitute model set for training the signal pattern calibration generative model, enabling the eavesdropping of user activity with an accuracy of up to 80%. Given that various hand coverage and finger movements create unique interference to signals, the authors in [28] extracted temporal and frequency domain fluctuations of CSI and employed dynamic time warping to achieve keystroke recognition, allowing eavesdrop on user passwords. Evaluations show that the system achieves an average classification accuracy of 93.5%.

The second type is to attack the legitimate sensing, leading the authorized system to make wrong decisions [11], as shown in part *c* of the Fig. 1. For example, the authors in [29] introduced a signal synthesis scheme to generate adversarial signals with specific motion features. Then it employed a closed-box attack strategy to address discrepancies between the perturbation space and the classifier's input space, misleading the sensing system into incorrect gesture classification. Experimental results demonstrate that the average attack success rate of this system is over 70%. The authors in [30] developed an adversarial perturbation to degrade sensing performance, while also designing a mask to confine the perturbation to targeted areas, minimizing its impact on communication. Evaluations indicate that it can reduce the accuracy of the sensing model to below 50%, while decreasing the impact of perturbation on bit error rate by 77.78%.

For above security concerns, researchers have developed some effective solutions [31]. For instance, in RF-protect [32], the authors designed a novel hardware reflector that can modify radio waves by creating fake reflections, thereby masking the reflection caused by the user and protects the user's privacy. In MIRAGE [33], the authors combined beamforming and phase adjustment to prevent signal receivers from accurately filtering out parameters of the direct signal from the user to the receiver. This prevents the unauthorized receivers from locating the user, thereby ensuring the security of the user's location information. Moreover, in WiShield [34], the authors designed a component at the receiver to change the amplitude and phase of the signals received by the antenna. This disrupts the signal parameter estimation, such as AoA, thereby preventing the unauthorized sensing of user location and activities.

As mentioned above, existing works have discussed security issue in sensing and proposed some solutions. However, these solutions often require additional equipment or prior knowledge, such as the location of unauthorized devices, which is impractical. Therefore, this paper introduces DFSS, which designs and generates safeguarding signals via diffusion model. Then, it modulates them onto pilot to mask fluctuations caused by user activities, preventing illegitimate sensing by unauthorized devices. Here, DFSS employs the random sampling based Markov diffusion model. Compared to non-Markov diffusion models [35], it offers greater diversity in the protective signals generation, making them difficult to replicate. Moreover, unlike above mentioned methods, DFSS does not require additional hardware or prior knowledge, enhancing its practicality in real-world scenarios.

III. System Design

A. System Overview

The system consists of three main modules, as shown in Fig. 2. First, by taking the spatial distribution of ISAC devices and user location as inputs, we propose D-CDM to generate a graph, which guides the activation of wireless links, as well as signal transmitters (Txs) and receivers (Rxs) for sensing. Next, we use a continuous conditional diffusion model to generate a safeguarding signal given the activated links, and locations of Txs, Rxs, and user. Finally, the Tx modulates the generated safeguarding signal onto the pilot and then transmits it for user sensing. In this way, the signal fluctuations caused by user activities are masked by the modulated safeguarding signal. Therefore, only authorized Rx can use the same diffusion model, inputs, and random seed to regenerate the safeguarding signal, thereby extracting true CSI data for effective sensing and communication. Unauthorized devices, however, are unable to reproduce the same signal for CSI extraction and user sensing, therefore shielding users from illegitimate surveillance.

B. Signal Model

Consider an ISAC link that includes an orthogonal frequency division multiplexing (OFDM) signal transmitter and a receiver. When the receiver receives signals, it estimates the CSI using the pilot shared between the transmitter and receiver for sensing and communication. Assuming there is no inter-carrier interference (ICI) [36], the pilot signals for the N subcarriers can be expressed as a diagonal matrix $\mathbf{X} = \mathrm{diag}\left(X[0], \ldots, X[N-1]\right)$, where $X[n]$ indicates the pilot signal on the n-th subcarrier. Let $H[n]$ be channel gain of the n-th subcarrier, then the received pilot signals can be expressed as $\mathbf{Y} \triangleq \mathbf{X} \times \mathbf{H}$, where $\mathbf{H} = [H[0], \ldots, H[N-1]]$. Based on $\mathbf{Y}$ and the predefined $\mathbf{X}$, $\mathbf{H}$ can be estimated by minimizing the following cost function

$$\begin{aligned} J\left(\hat{\mathbf{H}}\right) &= \left\|\mathbf{Y} - \mathbf{X}\hat{\mathbf{H}}\right\|^2 \\ &= \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{H}}\right)^{\mathrm{H}} \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{H}}\right) \\ &= \mathbf{Y}^H\mathbf{Y} - \mathbf{Y}^H\mathbf{X}\hat{\mathbf{H}} - \hat{\mathbf{H}}^H\mathbf{X}^H\mathbf{Y} + \hat{\mathbf{H}}^H\mathbf{X}^H\mathbf{X}\hat{\mathbf{H}}. \end{aligned} \quad (1)$$

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

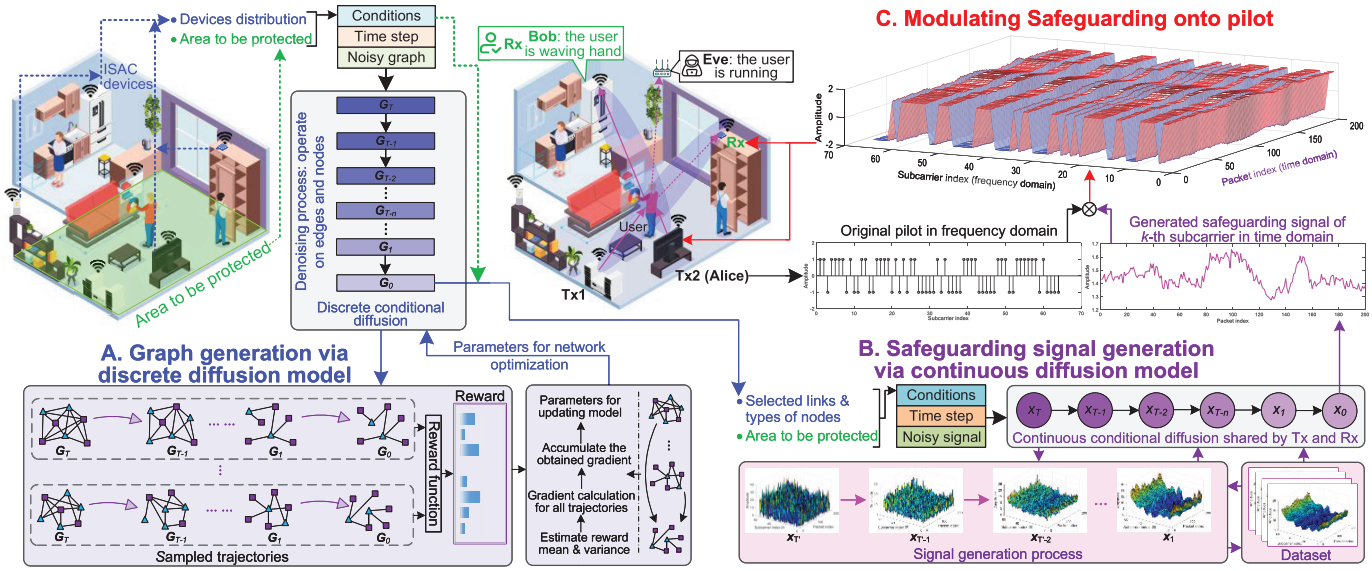


Fig. 2. The framework of the proposed DFSS. DFSS consists of three primary modules, including graph generation, safeguarding signal generation, and signal modulation. The first module generates the graph with nodes and edges based on the locations of users and the distribution of ISAC devices, guiding the activation of sensing links and nodes. Subsequently, the second module generates safeguarding signals by utilizing the obtained graphs and user locations. Finally, the third module modulates these safeguarding signals onto a pilot to mask the CSI amplitude fluctuations triggered by user activities. In this way, authorized devices can use the safeguarding signals to extract actual CSI from the captured signals for legitimate wireless sensing, whereas unauthorized devices are unable to do so, thereby preventing illegitimate sensing.

Let

$$\partial J\left(\hat{\mathbf{H}}\right)/\partial\hat{\mathbf{H}} = -2\left(\mathbf{X}^H\mathbf{Y}\right)^* + 2\left(\mathbf{X}^H\mathbf{X}\hat{\mathbf{H}}\right)^* = 0, \tag{2}$$

then the channel estimation can be obtained

$$\hat{\mathbf{H}} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{Y} = \mathbf{X}^{-1}\mathbf{Y}. \tag{3}$$

For the n-th subcarrier, the CSI at time t can be expressed as

$$H\left(f_n,t\right) = e^{-j\varepsilon}H_s\left(f_n,t\right) + e^{-j\varepsilon}\sum_{l\in P_d}\alpha_l\left(f_n,t\right)e^{-j2\pi f_n d_l(t)/c} + n_{f_n,t}, \tag{4}$$

where $e^{-j\varepsilon}$ is the phase offset, $H_s\left(f_n,t\right)$ is the sum of CSIs for static propagation paths, P_d is the set of dynamic paths caused by the moving user, $\alpha_l\left(f_n,t\right)$ is a complex number including the attenuation and initial phase of the l-th path, $e^{-j2\pi f_n d_l(t)/c}$ is the phase shift caused by the length change of the l-th dynamic propagation path, and $n_{f_n,t}$ is the noise.

Based on (4), we can obtain $|H\left(f_n,t\right)|^2$, shown as (5) at the bottom of the page. Here, $\Psi\left(n_{f_n,t}\right)$ represents the power of the cross terms that are multiplied by noise, v_l is the rate of length change of the l-th dynamic path, $d_l\left(0\right)$ is the initial length, θ_{sl} and $\theta_{ll'}$ denote initial phases, $\Delta v_{ll'} = v_l - v_{l'}$ is the rate difference between two paths, and $\Delta d_{ll'}\left(0\right) = d_l\left(0\right) - d_{l'}\left(0\right)$ represents the initial length difference. From (5), we can observe that the CSI power is composed of a series of constants and sinusoids, with the overall intensity dependent on $|H_s\left(f_n,t\right)|$. Additionally, the overall fluctuation characteristics of the CSI are primarily determined by the third term with the intensity smaller than $|H_s\left(f_n,t\right)|^2$, which is a constant. As different user activities result in various v_l, the user activity recognition can be achieved by extracting and analyzing fluctuation characteristics. To mitigate fluctuations caused by user behaviors, this paper proposes to generate the safeguarding signal $s\left(f_n,t\right)$, which is then multiplied with the pilot signals used for CSI estimation at the transmitter. Hence, based on $\mathbf{X}$, we have

$$\mathbf{X}' = \mathrm{diag}\left(s\left(f_0,t\right)X\left[0\right],\ldots,s\left(f_{N-1},t\right)X\left[N-1\right]\right). \tag{6}$$

The generated safeguarding signal is shared only between authorized transmitters and receivers. Therefore, based on (1), for the authorized receiver, we have $\hat{\mathbf{H}} = \left(\mathbf{X}'\right)^{-1}\mathbf{Y}'$, where $\mathbf{Y}'$ is the signal received when $\mathbf{X}'$ is transmitted. For the unauthorized receiver, $\hat{\mathbf{H}}' = \mathbf{X}^{-1}\mathbf{Y}'$. On this basis, according to (4) and (5), we have

$$\left|H'\left(f_n,t\right)\right|^2 = |s\left(f_n,t\right)|^2|H\left(f_n,t\right)|^2, \tag{7}$$

where

$$H'\left(f_n,t\right) = e^{-j\varepsilon}s\left(f_n,t\right)H_s\left(f_n,t\right) + {n'}_{f_n,t} + e^{-j\varepsilon}\sum_{l\in P_d}s\left(f_n,t\right)\alpha_l\left(f_n,t\right)e^{-j2\pi f_n d_l(t)/c}. \tag{8}$$

$$|H\left(f_n,t\right)|^2 = \sum_{l\in P_d}|\alpha_l\left(f_n,t\right)|^2 + |H_s\left(f_n,t\right)|^2 + \sum_{l\in P_d}2\left|H_s\left(f_n,t\right)\alpha_l\left(f_n,t\right)\right|\cos\left[2\pi f_n\left(v_l t + d_l\left(0\right)\right)/c + \theta_{sl}\right] + \sum_{l,l'\in P_d}\left(2\left|\alpha_l\left(f_n,t\right)\alpha_{l'}\left(f_n,t\right)\right| \times \cos\left(2\pi f_n\left(\Delta v_{ll'}t + \Delta d_{ll'}\left(0\right)\right)/c + \theta_{ll'}\right)\right) + \Psi\left(n_{f_n,t}\right). \tag{5}$$

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

For instance, if the CSI power is $|H_{wh}(f_n,t)|^2$ when the user waves hand, then the safeguarding signal can be set as $1/|H_{wh}(f_n,t)|$, thereby the CSI power received by unauthorized receivers is $1 = 1\Big/|H_{wh}(f_n,t)|^2 \times |H_{wh}(f_n,t)|^2$. In this way, the safeguarding signal can effectively mask signal fluctuation caused by user's gestures. Here, the fluctuation characteristics are determined by several parameters, mainly include P_d, v_l, and $d_l(0)$. While the overall intensity is primarily determined by the static paths, which are directly related to the locations of the signal transmitter and receiver. However, in practice, it is challenging to obtain these parameters, making it impossible to calculate $1/|H_{wh}(f_n,t)|$ directly.

Given the strong generation capabilities of GAI, hence, this paper first proposes D-CDM, which generates graph based on the given ISAC device distribution and user location to guide the activation of wireless links and signal Txs and Rxs. On this basis, the continuous diffusion model is utilized to generate the safeguarding signal by using the activated links, nodes, and user location as conditions. This signal is then modulated onto pilot signals to mask the signal fluctuations caused by user activities, thereby providing user with protection against unauthorized surveillance.

C. Graph Generation for Link Selection

Given the distribution of ISAC devices and the user location, we first determine the sensing link and signal Tx and Rx. Treating signal transmitters and receivers as nodes, sensing links as edges, and the user location and device distribution as conditions, this paper proposes D-CDM to generate a graph to guide the ISAC system to activate appropriate links and devices, ensuring effective and economical system operation.

1) Problem Formulation: Based on (6) to (8), we can observe that the dynamic path contains all the information related to user activity. Therefore, the sensing signal-to-noise ratio (SSNR), which is the ratio of the dynamic path power to the sum of the noise power and the static path power, can be used to characterize the sensing performance

$$SSNR = \frac{\sum_{l\in P_d} |\alpha_l(f_n,t)|^2}{\left|H_s(f_n,t) + n_{f_n,t}\right|^2}. \tag{9}$$

Here, a higher SSNR implies better sensing performance. According to [37], given the signal transceiver pair, the SSNR can be further expressed as

$$SSNR = \frac{\vartheta\sigma}{4\pi(d_{Tx}d_{Rx})^2\left(\gamma\vartheta\big/d_D^2 + b\right)}, \tag{10}$$

where ϑ is positively correlated with both signal transmission power and antenna gain, σ represents the effective reflection area of the user, d_{Tx} and d_{Rx} denote the distances between the user and the signal transmitter and receiver, respectively, γ and b are two constants for a given transceiver pair and can be measured in advance, and d_D is the path length of line-of-sight (LoS) path. As can be seen from (10), the sensing performance depends on the distance from the user to the signal transmitter and receiver. Given the spatial distribution of devices and user location, it is essential for an ISAC service provider to ensure the sensing performance while minimizing the cost. Thereby, the problem of selecting optimal sensing links, Txs, and Rxs can be formulated as the following optimization problem

$$\begin{aligned}
&\max_{\mathbf{G}}\{U_{LK} = \alpha_1 SSNR_T - \alpha_2 \#(Tx) - \alpha_3 \#(Link)\}\\
&s.t.,\ \#(\text{node}) > \#(Tx) \geq 1,\\
&\quad \#(\text{node}) > \#(Rx) \geq 1,\\
&\quad \#(Tx) \times \{\#(\text{node}) - \#(Tx)\} \geq \#(Link) \geq 1,\\
&\quad Tx \neq Rx,\\
&\quad Tx \nleftrightarrow T'x,\\
&\quad Rx \nleftrightarrow R'x,
\end{aligned} \tag{11}$$

where $SSNR_T$ is the sum of the SSNR for all links, $\mathbf{G}$ is the matrix that describes the link and node activation, α_1, α_2, and α_3 are weighting factors, $\#(Tx)$, $\#(Rx)$, $\#(Link)$, and $\#(node)$ denote the number of signal transmitters, receiver, links, and the total number of ISAC devices, respectively, $Tx \nleftrightarrow Tx'$ means no link among transmitters, and $Rx \nleftrightarrow Rx'$ indicates no link between receivers.[1]

2) Graph Generation: To solve the above optimization problem, we propose D-CDM to generate the optimal $\mathbf{G}_0$ for activation. Unlike diffusion models for images, which add and remove noise at each pixel, D-CDM operates on the nodes and edges [38]. Therefore, the state space consists of node types and edge types. Specifically, we adopt one-hot embedding [39] to encode node and edge types. This encoding is preferred primarily because it aligns with the discrete nature of the graph diffusion process, where transitions between categorical states are modeled via Markov matrices. Let k_i represent the one-hot encoding of the attributes for node i. Then, by organizing encodings into a matrix, we obtain $\mathbf{K}$. Similarly, $\mathbf{E}$ is used to group the one-hot encoding e_{ij} of each edge. On this basis, the optimal edge and node selection scheme can be expressed as $\mathbf{G}_0 = (\mathbf{K}_0, \mathbf{E}_0)$. For any node and edge, the transition probabilities at the t'-th step are defined as

$$\left[\mathbf{Q}_{t'}^K\right]_{ij} = q\left(k^{t'} = j \,\middle|\, k^{t'-1} = i\right) \tag{12}$$

and

$$\left[\mathbf{Q}_{t'}^E\right]_{i'j'} = q\left(e^{t'} = j' \,\middle|\, e^{t'-1} = i'\right), \tag{13}$$

respectively, where $i, j \in \{1,\ldots,a\}$ and $i', j' \in \{1,\ldots,a'\}$ are the types of nodes and edges, respectively. Based on these transition matrices, the forward diffusion process is

$$\begin{aligned}
q(\mathbf{G}_T|\mathbf{G}_0) &= \prod_{t'=1}^{T} q(\mathbf{G}_T|\mathbf{G}_{t'-1})\\
&= \prod_{t'=1}^{T} \left(\mathbf{K}_0\mathbf{Q}_{t'}^K, \mathbf{E}_0\mathbf{Q}_{t'}^E\right)\\
&= \left(\mathbf{K}_0\bar{\mathbf{Q}}_{t'}^K, \mathbf{E}_0\bar{\mathbf{Q}}_{t'}^K\right),
\end{aligned} \tag{14}$$

where T is the total number of noise addition steps.

Based on the forward diffusion, D-CDM performs conditional denoising to generate graphs using the denoising

[1]In the proposed framework, the Rx not only receives the wireless signal but also processes it, whereas the Tx only transmits the wireless signal, as typical communications. This implies that the Rx holds higher operational costs than that of the Tx. Hence, Tx and Rx are treated as two distinct types of nodes in our modeling.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

network parameterized by θ. Using the distribution of ISAC devices and user location as conditions and the noisy graph as input, the denoising network can generate a clear graph through denoising, which maximizes U_{LK} as defined in 11. Rigorously, This denoising process is

$$p_\theta\left(\mathbf{G}_0\left|\mathbf{G}_T\right., \mathbf{D}\right) = p\left(\mathbf{G}_T\right) \prod_{t'=1}^{T} p_\theta\left(\mathbf{G}_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right), \tag{15}$$

where $\mathbf{D}$ is the generation condition, and

$$p_\theta\left(\mathbf{G}_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right) = \underbrace{\prod_{1\le i\le a} p_\theta\left(k^i_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right)}_{node} \times \underbrace{\prod_{1\le i', j'\le a} p_\theta\left(e^{i'j'}_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right)}_{edge}. \tag{16}$$

For each term of a node, we have

$$\begin{aligned} &p_\theta\left(k^i_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right) \\ &= \int_{k^i} p_\theta\left(k^i_{t'-1}\left|k^i\right., \mathbf{G}_{t'}, \mathbf{D}\right) dp_\theta\left(k^i\left|\mathbf{G}_{t'}\right., \mathbf{D}\right) \\ &= \sum_{k\in A} p_\theta\left(k^i_{t'-1}\left|k^i = k, \mathbf{G}_{t'}, \mathbf{D}\right.\right) \hat{p}^i_K(k), \end{aligned} \tag{17}$$

where A is the space of categories for nodes.[2] Similarly, for each term of an edge, we have

$$\begin{aligned} &p_\theta\left(e^{i'j'}_{t'-1}\left|\mathbf{G}_{t'}\right., \mathbf{D}\right) \\ &= \sum_{e\in A'} p_\theta\left(e^{i'j'}_{t'-1}\left|e^{i'j'} = e, \mathbf{G}_{t'}, \mathbf{D}\right.\right) \hat{p}^{i'j'}_E(e), \end{aligned} \tag{18}$$

where A' is the space of categories for edges. Based on the constructed denoising network, we need to further refine its network parameters to facilitate the D-CDM in generating the optimal graph for sensing links and nodes activation. By using U_{LK} as the reward function, here the goal is to maximize the expected reward over the sample distribution

$$J_{ER}(\theta) = \mathbb{E}_{\mathbf{G}_0\sim p_\theta(\mathbf{G}_0)}\left[r\left(\mathbf{G}_0\right)\right], \tag{19}$$

where $r(\cdot)$ is the reward function defined by (11). However, optimizing $J_{ER}(\theta)$ directly is challenging as the likelihood $p_\theta(\mathbf{G}_0)$ is unavailable.

In DDPO [40], the authors considered the denoising process as a sequential decision operation and modeled it as a Markov process. Then, the policy gradient [41] can be derived and the diffusion model can be optimized using reinforcement learning (RL) methods. Specifically, at each timestep, the RL agent observes a state, takes an action according to the policy, obtains the reward, and then moves to the new state. As the agent acts in this Markov decision process, it produces trajectories, which are sequences of states and actions. The RL objective is for the agent to maximize the expected cumulative reward over trajectories sampled from its policy. Inspired by DDPO, we model the discrete graph diffusion as a Markov process:

$$\begin{cases} \mathbf{s}_{t'} \triangleq \left(\mathbf{G}_{T-t'}, T-t'\right) \\ \mathbf{a}_{t'} \triangleq \mathbf{G}_{T-t'-1} \\ \pi_\theta\left(\mathbf{a}_{t'}\left|\mathbf{s}_{t'}\right.\right) \triangleq p_\theta\left(\mathbf{G}_{T-t'-1}\left|\mathbf{G}_{T-t'}\right.\right) \\ r\left(\mathbf{s}_{t'}, \mathbf{a}_{t'}\right) \triangleq \begin{cases} r\left(\mathbf{G}_0\right), & t' = T \\ 0, & t' < T, \end{cases} \end{cases} \tag{20}$$

where $\mathbf{s}_{t'}$ and $\mathbf{a}_{t'}$ represent the state and action at the t'-th step, respectively, π_θ is the policy corresponding to the reverse transition distribution, and $r\left(\mathbf{s}_{t'}, \mathbf{a}_{t'}\right)$ is the reward. In this way, the graph generation process, denoted as $\{\mathbf{G}_T, \mathbf{G}_{T-1}, \ldots, \mathbf{G}_0\}$, can be regarded as a state-action trajectory produced by agent actions within the Markov process $\boldsymbol{\varsigma} = \{(\mathbf{s}_0, \mathbf{a}_0), (\mathbf{s}_1, \mathbf{a}_1), \ldots, (\mathbf{s}_T, \mathbf{a}_T)\}$. According to (20), we have $\sum_{t'=0}^{T} r\left(\mathbf{s}_{t'}, \mathbf{a}_{t'}\right) = r\left(\mathbf{G}_0\right)$. Hence, the expected reward of the agent can be denoted as

$$J_C(\theta) = \mathbb{E}_{p(\boldsymbol{\varsigma}|\pi_\theta)}\left[r\left(\boldsymbol{\varsigma}\right)\right] = \mathbb{E}_{p_\theta(\mathbf{G}_{0:T})}\left[r\left(\mathbf{G}_0\right)\right]. \tag{21}$$

We can obverse that $J_C(\theta)$ is equivalent to $J_{ER}(\theta)$. Since $J_C(\theta)$ is derived from the Markov process, we can use the RL to optimize it. On this basis, we further formulate the policy gradient corresponding to $J_C(\theta)$. Given $J_C(\theta) = \mathbb{E}_{p_\theta(\mathbf{G}_{0:T})}[r(\mathbf{G}_0)]$, we have

$$\nabla_\theta J_C(\theta) = \mathbb{E}_{\boldsymbol{\varsigma}}\left[r\left(\mathbf{G}_0\right) \sum_{t'=1}^{T} \nabla_\theta \log p_\theta\left(\mathbf{G}_{t'-1}\left|\mathbf{G}_{t'}\right.\right)\right], \tag{22}$$

which is generally intractable. Hence, we adopt Monte Carlo estimation to approximate (22), obtaining

$$\nabla_\theta J_C(\theta) \approx \frac{T\sum_{z\in Z}\sum_{t'\in\Gamma} r\left(\mathbf{G}_0^{(z)}\right) \nabla_\theta \log p_\theta\left(\mathbf{G}_{t'-1}^{(z)}\left|\mathbf{G}_{t'}^{(z)}\right.\right)}{|Z|\,|\Gamma|}, \tag{23}$$

where Z and Γ represent the sets of sampled trajectories and timesteps,[3] respectively, and $\mathbf{G}^{(z)}_{0:T}, z = 1, \ldots, Z$ are Z trajectories sampled from $p_\theta(\mathbf{G}_{0:T})$. Essentially, the policy gradient estimation in (23) is a weighted summation of gradients, which can produce fluctuating and unreliable policy gradient estimates when the number of Monte Carlo samples is limited. To address this issue, (23) is further modified into

$$\nabla_\theta J'_C(\theta) = \frac{T\sum_{z\in Z}\sum_{t'\in\Gamma} r\left(\mathbf{G}_0^{(z)}\right) \nabla_\theta \log p_\theta\left(\mathbf{G}_0^{(z)}\left|\mathbf{G}_{t'}^{(z)}\right.\right)}{|Z|\,|\Gamma|}. \tag{24}$$

As can be observed, the potential number of graph trajectories is vast. Therefore, we can group them into different equivalent classes where trajectories with the same $\mathbf{G}_0^{(z)}$ are considered equivalent. In this way, the number of these classes is less than the total number of graph trajectories, and the optimization can be performed over the classes, which is simpler than optimizing in the entire trajectory space. Finally, we perform gradient ascent to optimize the policy. Specifically, $\nabla_\theta J'_C(\theta)$ is used to update the network parameters via $\theta' = \theta + \nabla_\theta J'_C(\theta) \times \eta$, where η represents the learning rate. Up on completing the training, the denoising network can generate

[2]The categories of nodes should be defined according to a specific application. In this paper, nodes are mainly classified into two categories, including signal transmitters and signal receivers.

[3]Here, $|Z| = 256$, Γ is an integer randomly sampled from 1 to 30.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

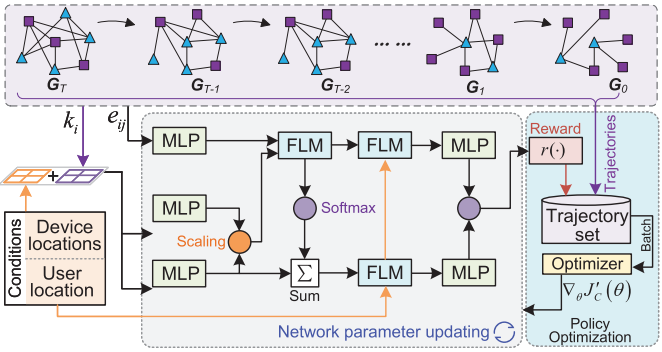


Fig. 3. The design of the denoising network and optimization process.

graphs based on new input conditions to guide the activation of links and nodes.

To facilitate learning of graph generation strategies by the denoising network while considering the generation conditions, we build the denoising network based on the graph transformer architecture [42], and the overall structure is demonstrated in Fig. 3. Specifically, the input nodes and edges are first processed by a multi-layer perceptron (MLP) module [43] to produce embeddings. These embeddings are then passed through a graph transformer layer that features an attention mechanism. Similar to conventional ones, the scoring at each attention layer is defined by feature-wise linear modulation, which can adjust the output of the transformer layer through scaling and shifting operation based on condition $\mathbf{D}$. Following this, the embeddings of nodes and edges undergo residual connections and layer normalization. These steps prevent the vanishing of gradients and ensure a stable training process. Finally, two separate MLPs are used to decode the node and edge embeddings, respectively, leading to the matrices corresponding to the prediction for $\mathbf{G}_0$.

D. Safeguarding Signal Generation

Using the generated $\mathbf{G}_z^0$, combined with user locations as conditions, we further train the continuous diffusion model to produce the safeguarding signal. Unlike the previous method, which optimizes the graph generation process using the reward function, here, to guarantee performance, we train the continuous diffusion model with the dataset containing safeguarding signals. This enables the model to generate the new safeguarding signal that is similar to yet distinct from the one used in training, thereby ensuring diversity and randomness.

1) Continuous Diffusion Model: Similar to discrete diffusion models, continuous conditional diffusion models also incorporate a forward process and a reverse process. Given a schedule of noise scales $0 < \beta_1, \ldots, \beta_{T'} < 1$, for any given safeguarding training sample $\mathbf{x}_0 \sim q'_{sg}(\mathbf{x})$, the forward process perturbs the training sample by adding the noise over T' steps, such that

$$q'(\mathbf{x}_{t'}|\mathbf{x}_{t'-1}) = \mathcal{N}\left(\mathbf{x}_{t'}; \sqrt{1-\beta_{t'}}\mathbf{x}_{t'-1}, \beta_{t'}\mathbf{I}\right), \tag{25}$$

where $\mathbf{I}$ is the identity matrix. Based on (25), we have

$$q'_{\alpha_{t'}}(\mathbf{x}_{t'}|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t'}; \sqrt{\alpha_{t'}}\mathbf{x}_0, (1-\alpha_{t'})\mathbf{I}\right), \tag{26}$$

where $\alpha_{t'} = \prod_{t''=1}^{t'}(1-\beta_{t''})$. After the addition of noise, the disrupted data distribution can be denoted as

$$q'_{\alpha_{t'}}(\hat{\mathbf{x}}) = \int q'_{sg}(\mathbf{x}) q'_{\alpha_{t'}}(\hat{\mathbf{x}}|\mathbf{x})\, d\mathbf{x}. \tag{27}$$

In the forward process described above, the noise $\beta_1, \ldots, \beta_{T'}$ is prescribed, allowing $\mathbf{x}_{T'}$ to approximate $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

The reverse process is defined as a Markov chain, which is parameterized by

$$p'_{\theta'}(\mathbf{x}_{t'-1}|\mathbf{x}_{t'}, \mathbf{D}') = \mathcal{N}\left(\mathbf{x}_{t'-1}; \frac{\mathbf{x}_{t'} + \beta_{t'}\boldsymbol{\mu}_{\theta'}(\mathbf{x}_{t'}, t', \mathbf{D}')}{\sqrt{1-\beta_{t'}}}, \beta_{t'}\mathbf{I}\right), \tag{28}$$

where $\mathbf{D}'$ is the generation conditions. The objective here is to train the denoising network to generate the safeguarding signal from the noise based on$\mathbf{D}'$, and ensure that the distribution of the generated signal is consistent with that of the training samples. Therefore, the denoising network is trained using a re-weighted version of the evidence lower bound shown in (29) at the bottom of the page. After training, the denoising network can start from $\mathbf{x}_{T'} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and generate safeguarding signal samples based on the generation conditions by following the reverse Markov chain, which can be recorded as

$$\mathbf{x}_{t'-1} = \frac{1}{\sqrt{1-\beta_{t'}}}\left[\mathbf{x}_{t'} + \beta_{t'}\boldsymbol{\mu}_{\theta''}(\mathbf{x}_{t'}, t', \mathbf{D}')\right] + \sqrt{\beta_{t'}}\mathbf{v}_{t'}, \tag{30}$$

where $t' = T', T'-1, \ldots, 1$ and $\mathbf{v}_{t'} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The safeguarding signals generated here are similar to, but distinct from those in the training dataset. Note that we use the random sampling based Markov diffusion model, rather than a non-Markov diffusion model. This is because the random sampling provides better diversity and randomness, preventing unauthorized Rxs from replicating the safeguarding signals.

2) Safeguarding Signal Design and Dataset Construction: To ensure the aforementioned continuous conditional diffusion model can generate effective safeguarding signals, it is crucial to construct a dataset to train the diffusion model. For instance, we can build a dataset that includes multiple $1/|H_{wh}(f_n, t)|$ for waving hand. Using this dataset, the diffusion model is trained to generate safeguarding signals, thereby protecting the user's gestures from unauthorized surveillance. However, in practice, an ISAC system cannot predict user's actions and generate corresponding safeguarding signals. Therefore, this paper designs a composite safeguarding signal and constructs the corresponding dataset to train the diffusion model.

$$\theta'' = \arg\min_{\theta'} \sum_{t'=1}^{T'} (1-\alpha_{t'}) \mathbb{E}_{q'_{data}(\mathbf{x})} \mathbb{E}_{q'_{\alpha_{t'}}(\hat{\mathbf{x}}|\mathbf{x})} \left[\left\| \boldsymbol{\mu}_{\theta'}(\hat{\mathbf{x}}, t', \mathbf{D}') - \nabla_{\hat{\mathbf{x}}} \log q'_{\alpha_{t'}}(\hat{\mathbf{x}}|\mathbf{x}) \right\|_2^2 \right]. \tag{29}$$

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

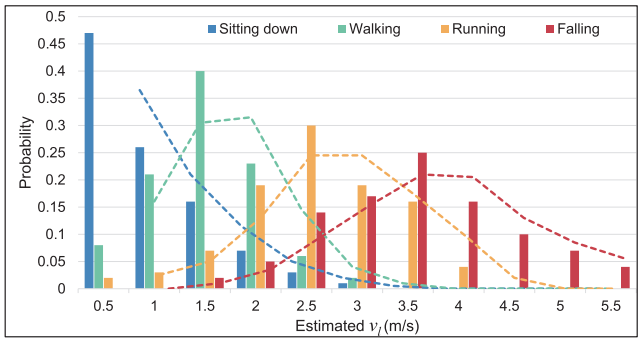


Fig. 4. The v_l of different activities in typical indoor scenario.

Specifically, in the absence of user activity, the environment contains only static paths with $v_l = 0$. In this scenario, both the received signal and the extracted CSI are relatively stable. When user activity occurs, it introduces more dynamic reflections characterized by $v_l \neq 0$. For these dynamic reflections, $v_l \neq 0$ leads to fluctuations in various signal parameters, such as path attenuation and phase shift. These fluctuations are recorded in the CSI, as described in (4). Therefore, $v_l \neq 0$ is a primary cause of signal fluctuations. Existing research demonstrates that different user activities cause the distinct rate of length change v_l. We conducted experiments in a typical indoor scenario based on the method described in [44], and performed statistical analysis on v_l corresponding to four typical user activities. The results are shown in Fig. 4. As can be seen, for the sitting down, the range of v_l extends from approximately 0.2 m/s to 1 m/s, while for falling, it ranges from about 0.8 m/s to 1.5 m/s, which is greater than the other three activities. Given the range of v_l caused by indoor activities is limited (approximately 0 to 5.5 m/s) and the purpose of the safeguarding signal is to mask the signal fluctuations caused by user activities, this paper constructs a dataset with multiple composite safeguarding signals for training the continuous diffusion model. Here, each composite safeguarding signal is defined as

$$s(f_n,t) = 1/|H_{st}(f_n,t)| + 1/|H_{wk}(f_n,t)| + 1/|H_{rn}(f_n,t)| + 1\Big/\left|H_{fl}(f_n,t)\right|, \tag{31}$$

where $1/|H_{st}(f_n,t)|$, $1/|H_{wk}(f_n,t)|$, $1/|H_{rn}(f_n,t)|$, and $1\Big/\left|H_{fl}(f_n,t)\right|$ are the safeguarding signal corresponding to siting down, walking, running and falling, respectively.

The reason for selecting these activities to construct the composite safeguarding signal is that v_l corresponding to these activities ranges from approximately 0 to 5.5 m/s [45], covering the almost all possible ranges in indoor scenarios. Therefore, using $s(f_n,t)$ yields two distinct advantages: a) $s(f_n,t)$ can effectively cancel signal fluctuations caused by most user activities in indoor scenarios; b) $s(f_n,t)$ can introduce certain interference, further masking the fluctuation characteristics. For example, for waving hand, according to the possible range of v_l, the components $1/|H_{st}(f_n,t)|$ and $1/|H_{wk}(f_n,t)|$ in $s(f_n,t)$ can partially cancel the signal fluctuations. At the same time, $1/|H_{rn}(f_n,t)|$ and $1\Big/\left|H_{fl}(f_n,t)\right|$ can introduce additional interference, thereby further obscuring the signal fluctuation characteristics. To construct the dataset for training, we utilize the Universal Software Radio Peripheral (USRP) N321 to collect the CSI corresponding to the four types of user activities under various conditions. Subsequently, we compute $1/|H_{st}(f_n,t)|$, $1/|H_{wk}(f_n,t)|$, $1/|H_{rn}(f_n,t)|$, and $1\Big/\left|H_{fl}(f_n,t)\right|$, and then sum them to obtain the composite safeguarding signal as expressed in (31). Figure 5 illustrates the dataset construction process.

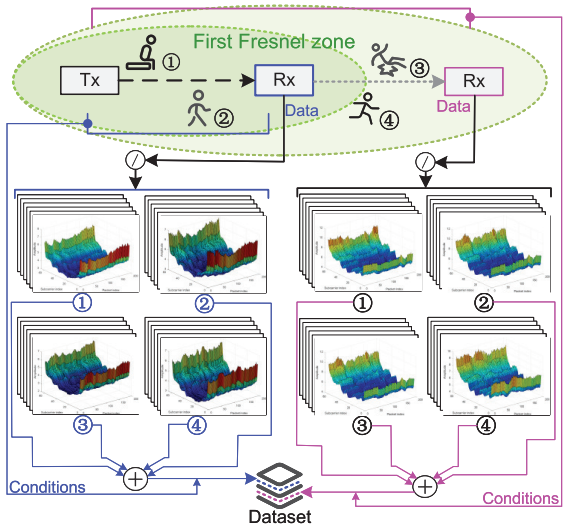


Fig. 5. The dataset construction process. Based on the CSI corresponding to four types of activities collected under various conditions, we calculate s via (31). Using conditions as labels, we aggregate multiple computed s values to construct the training dataset.

E. Safeguarding Signal Modulation

Based on the constructed dataset, the continuous diffusion model is trained to generate composite safeguarding signals under the given conditions during operation. These generated signals possess the same distribution as those in the dataset and exhibit certain diversity and randomness, thereby preventing unauthorized receivers from replicating this process to generate identical signals. Then, the generated signals are modulated onto the real part of the pilot signal. Let the generated protected signal be

$$\mathbf{s} = \left[s(f_n,1),\ldots,s(f_n,w),\ldots,s(f_n,W)\right], \tag{32}$$

where W is the length of the safeguarding signal, and the original pilot signal is $\pm 1 + 0\mathrm{i}$, then the pilot signal in the data packet transmitted at time w is $\pm 1 \times s(f_n,w) + 0\mathrm{i}$, so that

$$\mathbf{X}'_w = \mathrm{diag}\left(s(f_n,w)\,X[0],\ldots,s(f_n,w)\,X[N-1]\right). \tag{33}$$

This indicates that the safeguarding signals assigned to pilots at different frequencies are the same at time w. In this way, the amplitude of the pilot signal in the time domain varies in accordance with the trend represented by the generated safeguarding signal, thereby masking the fluctuation characteristics caused by user activities. Leveraging the same model and random seed, authorized Rx can generate the same safeguarding signal as the Tx and multiply it with the original pilot signal to obtain $\mathbf{X}'$. This allows for effective estimation of

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

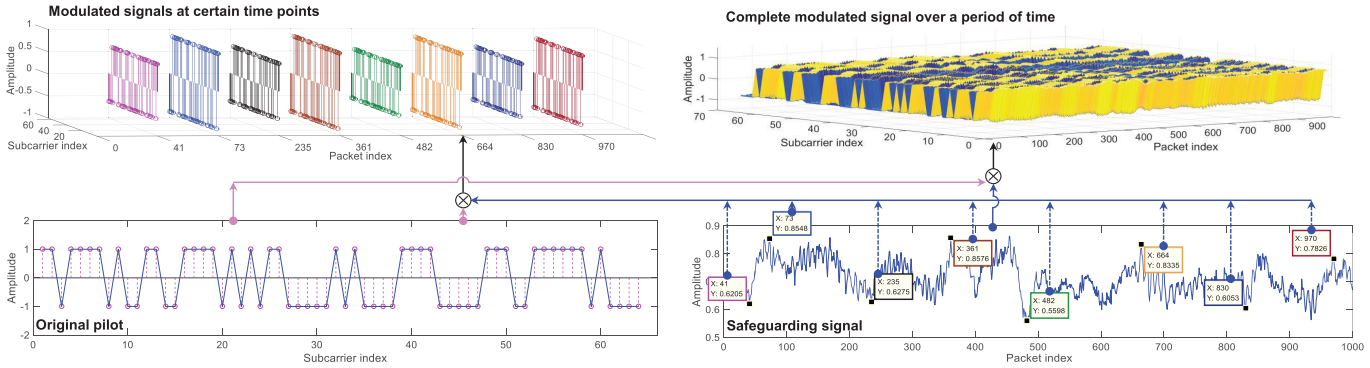


Fig. 6. The process of modulating the safeguarding signal onto the original pilot. Here, we show how safeguarding signals are modulated at some certain time points and over a period of time.

$\hat{\mathbf{H}}$ according to (1)–(3). However, unauthorized receivers can only estimate channels using the original pilot signal, resulting in $\hat{\mathbf{H}}' = \mathbf{X}^{-1}\mathbf{Y}'$, which prevents them from illegitimate sensing. Figure 6 shows the process of modulating the generated signal onto the pilot. Note that this operation leaves the phase of the CSI unchanged, which means that it does not influence the estimation of parameters such the AoA or the associated applications.

F. Complexity Analysis

We focus on the complexity of two diffusion models in DFSS. First, the discrete graph diffusion model D-CDM is used to generate the optimal links. Suppose that the graph has N_d nodes (with N_d^2 edges). E denotes the number of training episodes, D_t denotes the number of trajectories collected in one episode, T represents the number of diffusion steps, L denotes the number of Transformer layers in D-CDM's denoising network, B_{sz} denotes the batch size, and $|\theta|$ denotes the parameter size of the denoising network. For each training episode, the complexity to initialize a random graph is $\mathcal{O}(N_d^2)$. Then, the graph generation process has a complexity of $\mathcal{O}\left(D_t T L N_d^2\right)$. This is because the calculation of attention causes a complexity of $\mathcal{O}(N_d^2)$ in each layer [42]. Finally, the parameter update consumes $\mathcal{O}(B_{sz}|\theta|)$. Hence, the overall computational complexity of D-CDM training can be expressed as $\mathcal{O}\Big(E \times \left(D_t\left(T L+1\right) N_d^2 + B_{sz}|\theta|\right)\Big)$. The corresponding inference computational complexity is $\mathcal{O}\Big(\left(T L+1\right) N_d^2\Big)$. For the continuous diffusion model to generate safeguard signals, the training complexity can be expressed as $\mathcal{O}(N_t|\theta'|)$, where N_t and $|\theta'|$ denote the number of training samples and denoising network parameters, respectively. The corresponding inference complexity is $\mathcal{O}(T'|\theta'|)$, where T' denotes the number of diffusion steps [17]. It can be seen that for both discrete and continuous diffusion models, the complexity increases with an increase in the number of denoising steps. Hence, further research on accelerated sampling techniques to reduce the required denoising steps would be valuable for improving the efficiency of DFSS.

IV. Implementation and Evaluation

In this section, we utilize the USRP to evaluate the proposed method, covering three aspects. First, we analyze the graph generation for the link and nodes activation. Next, we evaluate the safeguarding signal generation and modulation. Finally, the performance in protecting users from unauthorized monitoring is assessed, using human activity recognition as a case study.

A. Experimental Design

1) Experimental Configurations: Based on the server and USRP N321 device, we conduct experiments in two different typical indoor scenarios. Specifically, the training and inference of the diffusion model are performed on a server running the Ubuntu 20.04 operating system, equipped with the AMD Ryzen Threadripper PRO 3975WX 32-core processor and the NVIDIA RTX A5000 GPU. The generated signals are then transmitted to another server, which has 64 GB of memory and GNU Radio,[4] and connects to USRP devices via optical fibers and 10-gigabit Ethernet cables. Each N321 device[5] includes the Xilinx Zynq-7100 SoC based baseband processor and UBX-160 daughterboard,[6] supporting up to 200 MHz bandwidth. The devices are synchronized using the OctoClock-G CDA-2990.[7] During the experiments, the signal center frequency is set at 2.8 GHz, with a bandwidth of 100 MHz covered by 64 subcarriers. We adopt a block-type pilot arrangement and transmit data at the rate of 100 packets per second. Additionally, to ensure the quality of the training data, the directional antenna with a gain of 12 dBi is used for transmitting and receiving signals. Figure 7 displays the prototype diagram and part of the GNU projects. The experimental scenarios includes an office and a meeting room, with the size of approximately 12.5m $\times$10m and 9m $\times$9m, respectively. Each room is equipped with multiple rows of tables and chairs, along with other pieces of furniture made of wood and metal, creating a indoor environment with rich multipath signal. User activities such as walking and sitting occur frequently in these spaces, making them key areas for wireless sensing. The detailed layouts of these rooms are shown in Fig. 8.

2) Experimental Method: We first train the proposed D-CDM, where the input includes device locations and user location, and the output is the generated graph depicting the

[4]https://www.gnuradio.org/
[5]https://www.ettus.com/all-products/usrp-n321/
[6]https://www.ettus.com/all-products/ubx160/
[7]https://www.ettus.com/all-products/octoclock-g/

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

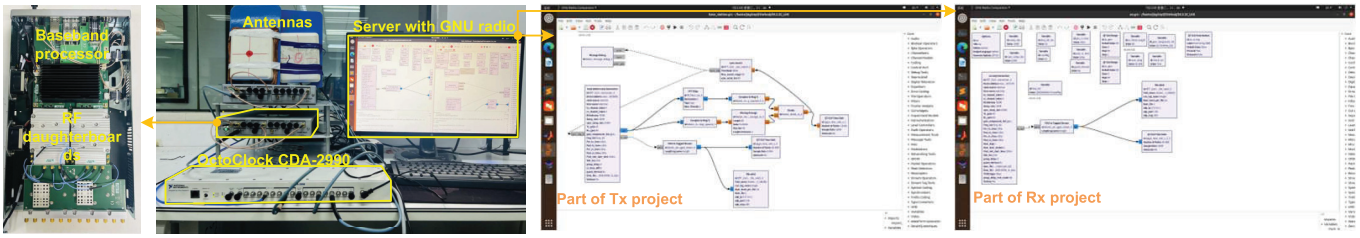


Fig. 7. The hardware equipment and GNUradio projects for experiments. Here, we display the server, USRP N321 (and its internal structure), an external clock, and the used directional antenna. Certain parts of the GNUradio project for Tx and Rx are also presented.

Fig. 8. The experimental scenarios. The original training data is collected in the office room and the evaluations are conducted in both rooms.

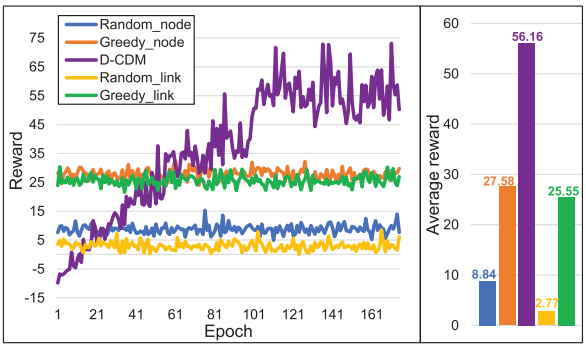


Fig. 9. The reward of D-DCM and other baselines over training epochs. Note that the average reward is calculated by averaging 50 testing rewards after the convergence.

wireless links, Tx, and Rx that need to be activated. This generated graph is evaluated by reward function, and the feedback is used to optimize the network parameters to complete the training. The continuous diffusion model is then trained with the dataset that contains composite safeguarding signals. The dataset comprises 20,000 samples, each constructed using the method described in (31). Specifically, we first collect CSI for four types of activities under various conditions in the office room, with each activity including 1,000 samples. Subsequently, for the same conditions, we randomly combine the CSI data of different activities to compute the composite protected signals. Finally, the computed sample is linked with the corresponding conditions, i.e., the locations of the Tx, Rx, and the user, to produce the final training sample.

During the evaluation, we first put the spatial distribution of ISAC devices and user location into the trained discrete diffusion model to analyze the graph generation performance, including the generation process and final outcomes. After that, the activated links, nodes, and user location are fed to continuous diffusion model as conditions to analyze the safeguarding signal generation. Finally, we modulate the generated signals onto the pilot and use the activity recognition methods in [46], [25], [47], and [48] to evaluate the system performance of DFSS in protecting users from unauthorized sensing. Moreover, using the trained model, we perform the same activity recognition evaluation in a meeting room, so as to analyze the robustness of the DFSS.

3) Evaluation Metrics: For the generated safeguarding signals, we use the structural similarity index measure (SSIM) and Fréchet inception distance (FID) to evaluate the similarity between the generated signals and those in the training dataset. SSIM captures local patterns of structural information between the generated and training signals, producing a score from 0 (no similarity) to 1 (perfect similarity) to measure how closely the generated signal aligns with the training signal [16]. SSIM is particularly useful for capturing distortions in spatial structure, which is critical in safeguarding signal generation where preserving fine-grained details is important. FID measures how closely the distributions of features, extracted by the deep neural network, match between real and generated signals. It calculates the Fréchet distance between the two sets of feature embeddings, capturing high-level similarities. A lower FID indicates that the generated signals and real signals are statistically more similar. By combining SSIM's focus on local patterns and FID's emphasis on overall distributional alignment, we can gain a comprehensive view of the quality of the generated signals. For activity recognition, the recognition accuracy (RA) and the accuracy degradation rate (ADR) are used to evaluate the system performance. Here, RA is calculated by dividing the number of correct recognitions by the total number of tests. The ADR is defined as $ADR = \left(Ac_{org} - Ac_{sf}\right)/Ac_{org}$, where Ac_{org} is the RA without using safeguarding signal, and Ac_{sf} is the RA of unauthorized APs when safeguarding signals are used. A higher ADR means better protection performance.

B. Experimental Results

1) Graph Generation: First, we analyze the graph generation performance of the proposed D-CDM. Figure 9 shows the

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

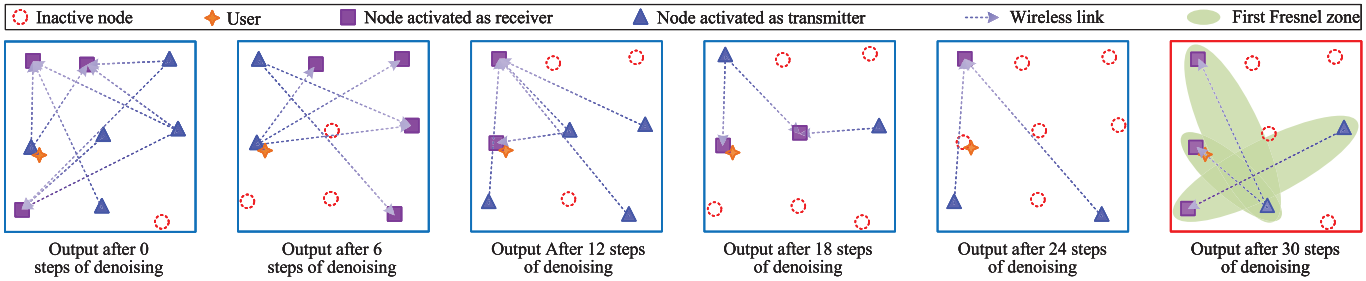


Fig. 10. The graph generation process based on trained D-CDM.

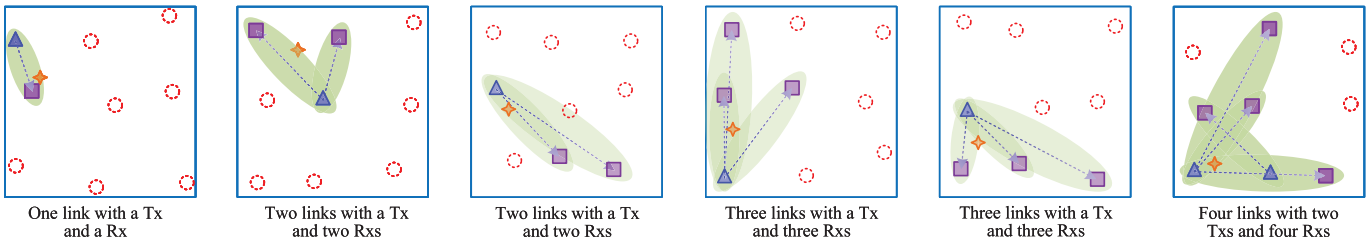


Fig. 11. The graph generation results under different generative conditions, i.e., different ISAC device distributions.

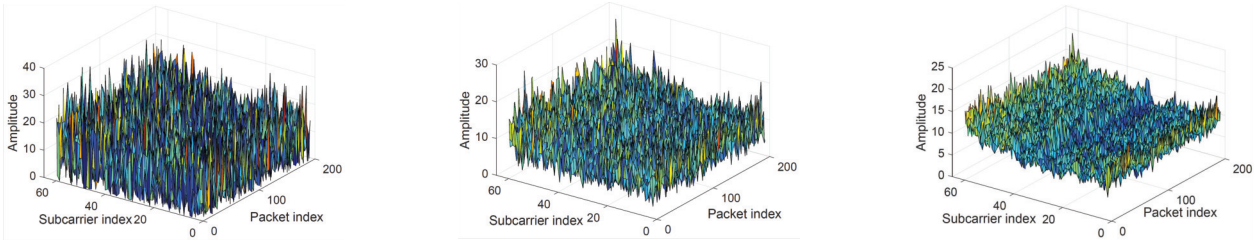
training curve of D-CDM and compares it with four baselines. Here, node-based and link-based greedy methods activate the four nodes and links closest to the user, respectively. Hence, the graph can be generated based on the activated parts (for Greedy-node, the activated nodes and the associated links are included; for Greedy-edge, the activated links and the connected nodes are included). The Tx/Rx roles of nodes are assigned following the rules defined in (11). In contrast, random methods activate nodes and links randomly. As can be seen, D-CDM converges after roughly 100 epochs with an average reward of about 56.16, indicating effective network parameter optimization and generation strategy learning via feedback from the reward function. In contrast, the average rewards for node-based and link-based random methods are approximately 8.84 and 2.77, respectively, and greedy approaches yield rewards of 27.58 and 25.55, respectively. This can be interpreted by that D-CDM generates graphs from random noise, introducing a degree of randomness. This randomness endows D-CDM with stronger exploratory capabilities, allowing it to find solutions with better rewards. Moreover, the D-CDM considers the impact of both nodes and edges during the denoising process, making the denoising more comprehensive and generating graphs that better meet the requirements.

Second, Fig. 10 depicts the graph generation process by the trained D-CDM under various input conditions. As presented, when the number of steps is less than 12, the denoising is incomplete, resulting in graphs with noisy edges and nodes, which are less useful for effective sensing. As the denoising steps increase, D-CDM continuously refines the noisy graph by appropriately adding or removing nodes and edges. After 18 steps, there is a noticeable decrease in the number of ineffective edges and nodes. Ultimately, an optimal graph is produced after 30 denoising steps, as indicated by the results in the red box. In this graph, the activated links form several Fresnel zones around the user, allowing the ISAC system to sense more efficiently. Figure 11 displays the generated results under different spatial distributions of ISAC devices. The generated graphs indicate that the trained D-CDM can adjust the node and edge generations according to the input conditions, hence generating the desired graphs under various conditions, demonstrating the adaptability of the proposed D-CDM. In addition, the results show that when the generated optimal graph has 4 wireless links, it contains 2 Txs and 4 Rxs, which is more energy-efficient than 3 Txs and 4 Rxs, as using more Tx consumes more energy. This further validates the rationality of D-CDM.

2) Safeguarding Signal Generation: Using the activated links, nodes, and user location as generation conditions, we further analyze safeguarding signal generation. Figures 12 and 13 display the generation process under different conditions. As shown, starting with random noise, when the number of steps is less than 460, the generated signals are noisy, indicating insufficient denoising. As denoising continues, noise is further removed, making the safeguarding signals clearer and more structured. After completing 500 steps, we obtain the generated safeguarding signals, which exhibit similar trends to the training data but are distinct from that, confirming the effectiveness of the safeguarding signal generation.

Additionally, different input conditions yield unique safeguarding signals, particularly in signal strength and trend. For instance, when the spatial positions of the Tx and Rx are close in the generation conditions, the resulting safeguarding signals are weak, whereas they are stronger when Tx and Rx are further apart. This happens because the Rx captures signals with higher amplitude and clear fluctuation characteristics when Tx and Rx are near each other, necessitating weaker safeguarding signals to effectively mask fluctuations caused by user activities. Conversely, a greater distance between Tx and Rx demands a relatively stronger safeguarding signal. Building

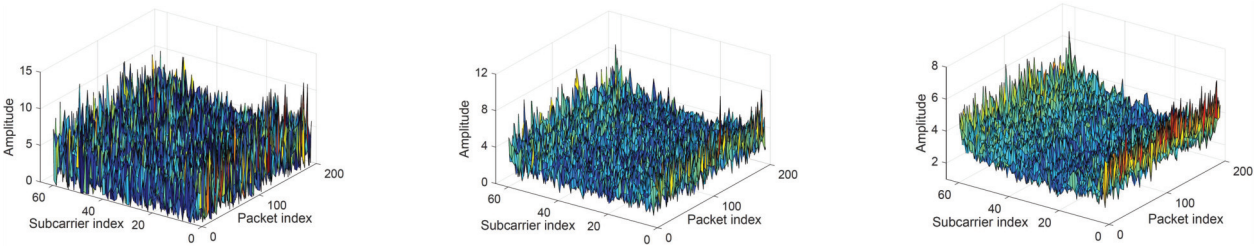
Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.



(a) The result after 400 steps of denoising. (b) The result after 460 steps of denoising. (c) The result after 480 steps of denoising.

(d) The result after 490 steps of denoising. (e) The result after 500 steps of denoising. (f) The training data from the dataset.

Fig. 12. The safeguarding signal generation process when Tx and Rx are about 6 meters apart.



(a) The result after 400 steps of denoising. (b) The result after 460 steps of denoising. (c) The result after 480 steps of denoising.

(d) The result after 490 steps of denoising. (e) The result after 500 steps of denoising. (f) The training data from the dataset.

Fig. 13. The safeguarding signal generation process when Tx and Rx are about 3 meters apart.

on this, we compare the generated safeguarding signals with the training data. The results show that while the generated signals closely match the training data in intensity and overall trends, they remain distinct. This highlights the diversity and randomness of the signals produced by the diffusion model.

Subsequently, we evaluate the quality of the generated signals using the FID and SSIM metrics, and compare the proposed DFSS with other two methods. The results, presented in Fig. 14, show that DFSS achieves a median SSIM of 0.73. In comparison, methods based on GAN [49] and VAE [50] achieve median SSIMs of 0.68 and 0.57, respectively. Additionally, the median FID of DFSS is 3.5, which is lower than 5.8 and 7.1 of GAN-based and VAE-based methods, respectively. Such performance of DFSS can be attributed to two main factors. First, the diffusion model in DFSS features strong exploratory capabilities and a progressive denoising approach. This allows for thorough exploration and helps retain more detailed features of the signal during the generation process. Second, DFSS takes into account the generative

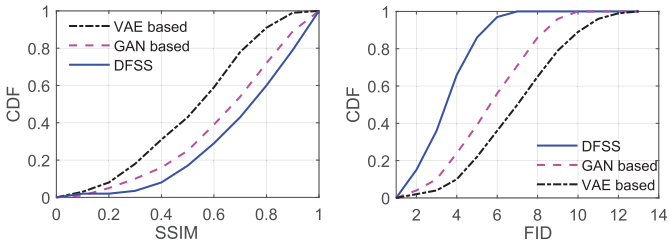
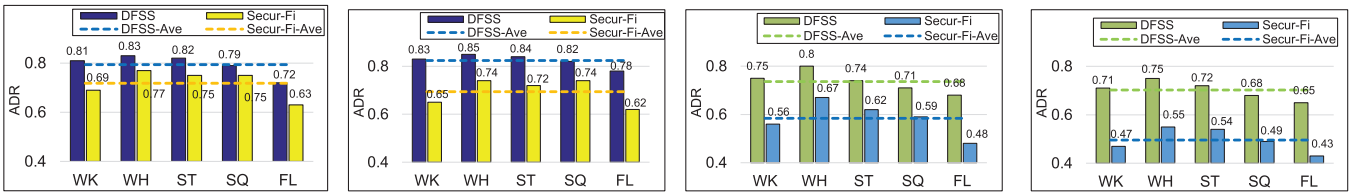


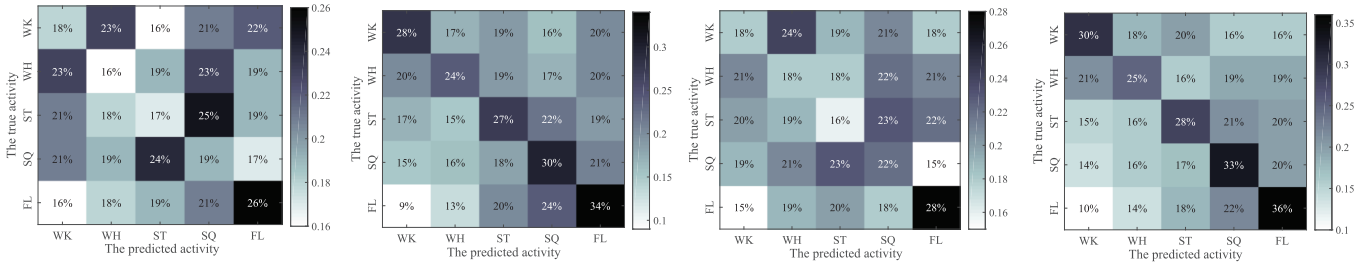
Fig. 14. The quality of the generated safeguarding signal.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.



(a) AF-ACT. (b) ABLSTM. (c) PhaseAnti. (d) CeHAR.

Fig. 15. The impact of DFSS on user activity recognition for unauthorized devices.



(a) AF-ACT. (b) CeHAR. (c) AF-ACT. (d) CeHAR.

Fig. 16. The recognition accuracy analysis of AF-ACT and CeHAR. (a) and (b) show the results in an office room, while (c) and (d) show the results in a meeting room.

conditions during both training and inference. This enables it to produce safeguarding signals that are better tailored to the specified conditions.

3) Protection Performance in Activity Recognition: Based on the activity recognition methods introduced in [46], [25], [47], and [48] (denoted as AF-ACT, ABLSTM, PhaseAnti, and CeHAR, respectively), we analyze the impact of DFSS on recognition accuracy using 5 actions (including walking (WK), waving hand (WH), sitting (ST), squatting (SQ), and falling (FL)) and compare it with Secur-Fi [31]. As shown in Fig. 15, with DFSS, the average ADR of the these systems can reach 0.82, 0.79, 0.74, and 0.7, respectively. These are higher compared to those achieved by Secur-Fi, which are 0.69, 0.72, 0.58, and 0.50, respectively. This is because DFSS modulates the generated safeguarding signal onto the pilot, directly influencing signal fluctuations. Compared with the Secur-Fi, which introduces interference by adjusting the antenna, our approach affects the signal more thoroughly and directly, thereby achieving better performance.

Additionally, DFSS is effective for different activities. For instance, in AF-ACT, the ADR for the five activities are 0.81, 0.83, 0.82, 0.79, and 0.72, respectively, demonstrating DFSS's general applicability across different activities. From the system perspective, the impact of DFSS varies. Specifically, DFSS has a more significant impact on AF-ACT and ABLSTM compared with PhaseAnti and CeHAR. This is because the former two rely solely on CSI amplitude for recognition, while the latter two use both amplitude and phase information, enhancing their robustness. Overall, DFSS effectively declines the performance of unauthorized systems, thereby achieving the goal of user protection.

Using AF-ACT and CeHAR as examples, Figs. 16(a) and (b) show the activity recognition confusion matrices with DFSS applied. Specifically, for five activities, AF-ACT's recognition accuracies are 18%, 16%, 17%, 19%, and 26%, while those of CeHAR are 28%, 24%, 27%, 30%, and 34%, respectively. This further demonstrates the robust protective capability of DFSS. Importantly, the results indicate that DFSS does not lead the systems to incorrectly classify different activities into the same one. This happens because the generated safeguarding signals are diverse and random, causing the CSI features captured by unauthorized devices to vary, which ultimately results in unpredictable recognition outcomes.

On this basis, we conduct cross-scenario experiments in a meeting room using the trained model, with results illustrated in Figs. 16(c) and (d). As can be seen, the confusion matrix reveals a slight performance decline. For instance, the recognition accuracy of falling in meeting room is improved by 2% for both AF-ACT and CeHAR. Such fluctuation is understandable given the differences between meeting room and office, as well as the original training data sourced from the office environment. Overall, in the meeting room, the average recognition accuracies of AF-ACT and CeHAR are 20.4% and 30.4%, respectively, which are comparable to those in the office. This demonstrates that DFSS can still provide effective protection, demonstrating its robustness.

Finally, we explored the impact of communication speed (i.e., data packet transmission rate) on DFSS. The results in Fig. 17 reveal that across different transmission rates, the average ADR of each system remains nearly consistent. For example, with data packet transmission rate ranging from 100 to 600, ADRs for AF-ACT are 0.79, 0.78, 0.80, 0.81 0.80, and 0.8, and for CeHAR, they are 0.70, 0.71, 0.69, 0.71, 0.70, and 0.72. This is because the generated safeguarding signals can be resampled to adapt to different communication speeds, ensuring DFSS's performance under varying conditions, which is critical for practical applications.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

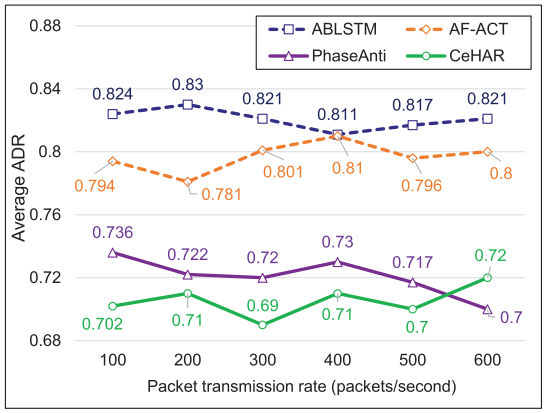


Fig. 17. The impact of communication speed on DFSS.

C. Limitation Analysis

The above evaluation results validate the effectiveness of the DFSS. However, it is undeniable that DFSS has certain limitations, which include the following aspects.

- Training stable and reliable diffusion models requires rich and high-quality original data. However, in practical ISAC scenarios, obtaining such raw data may necessitate specialized equipment and professional operators, which can be costly.
- Running diffusion models in DFSS is time-consuming and resource-intensive when many denoising steps are required. Therefore, DFSS faces practical challenges with resource consumption and inference latency.
- DFSS focuses on pilot amplitude, which does not affect the estimation of signal parameters such as ToA and AoA, which rely on CSI phase. However, it affects parameters like RSSI. Therefore, systems that use this kind of parameters have to utilize the generated safeguarding signal to parse the actual value, thereby mitigating this impact.

V. Conclusion

We have proposed DFSS, a secure sensing system that generates safeguarding signals and modulates them onto the pilot to mask signal fluctuations caused by user activity, thereby protecting users from unauthorized monitoring. In DFSS, we have developed a discrete conditional diffusion model to generate graphs that guide the ISAC system in activating appropriate links and nodes, ensuring economical operation while maintaining sensing performance. Additionally, we have designed a composite safeguarding signals for typical indoor activities and constructed a dataset to train the continuous conditional diffusion model, enabling it to generate safeguarding signals based on the input conditions. These signals, characterized by diversity and randomness, can effectively mask signal fluctuations features due to user activities, thereby preventing unauthorized devices from extracting actual CSI for illegitimate monitoring. Using activity recognition as an example, the evaluation shows that DFSS can reduce the recognition accuracy of unauthorized devices by approximately 70%, demonstrating the potential of GAI in enhancing sensing security. Future work will explore the use of GAI for secure user localization and tracking.

References

[1] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE Netw.*, vol. 35, no. 5, pp. 158–167, Sep. 2021.

[2] Z. Sun, G. Sun, Y. Liu, J. Wang, and D. Cao, "BARGAIN-MATCH: A game theoretical approach for resource allocation and task offloading in vehicular edge computing networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1655–1673, Feb. 2023.

[3] J. Zheng, J. Zhang, and B. Ai, "UAV communications with WPT-aided cell-free massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3114–3128, Oct. 2021.

[4] J. Li et al., "Active RIS-aided NOMA-enabled space- air-ground integrated networks with cognitive radio," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 314–333, Jan. 2025.

[5] Y. Zhang, L. Yang, X. Li, K. Guo, and H. Liu, "Covert communications for STAR-RIS-assisted industrial networks with a full duplex receiver and RSMA," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22483–22493, Dec. 2024.

[6] R. Kong and H. Chen, "CSI-RFF: Leveraging micro-signals on CSI for RF fingerprinting of commodity WiFi," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 5301–5315, 2024.

[7] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2019.

[8] J. Wang et al., "A unified framework for guiding generative AI with wireless perception in resource constrained mobile edge networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 11, pp. 10344–10360, Nov. 2024.

[9] D. Avola, M. Cascio, L. Cinque, A. Fagioli, and C. Petrioli, "Person re-identification through Wi-Fi extracted radio biometric signatures," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1145–1158, 2022.

[10] J. Wang, H. Du, D. Niyato, M. Zhou, J. Kang, and H. Vincent Poor, "Acceleration estimation of signal propagation path length changes for wireless sensing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11476–11492, Sep. 2024.

[11] J. Liu, Y. He, C. Xiao, J. Han, and K. Ren, "Time to think the security of WiFi-based behavior recognition systems," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 1, pp. 449–462, Feb. 2023.

[12] Y. Lin et al., "Blockchain-based efficient and trustworthy AIGC services in metaverse," *IEEE Trans. Services Comput.*, vol. 17, no. 5, pp. 2067–2079, Oct. 2024.

[13] J. Wang et al., "Generative AI for integrated sensing and communication: Insights from the physical layer perspective," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 246–255, Oct. 2024.

[14] Y. Lin et al., "A unified framework for integrating semantic communication and AI-generated content in metaverse," *IEEE Netw.*, vol. 38, no. 4, pp. 174–181, Jul. 2024.

[15] J. Wang et al., "Generative artificial intelligence assisted wireless sensing: Human flow detection in practical communication environments," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2737–2753, Oct. 2024.

[16] G. Chi et al., "RF-diffusion: Radio signal generation via time-frequency diffusion," 2024, *arXiv:2404.09140*.

[17] H. Du et al., "Enhancing deep reinforcement learning: A tutorial on generative diffusion models in network optimization," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2611–2646, 4th Quart., 2024.

[18] J. Wang et al., "Through the wall detection and localization of autonomous mobile device in indoor scenario," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 1, pp. 161–176, Jan. 2024.

[19] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316–327, Jan. 2020.

[20] H. F. T. Ahmed, H. Ahmad, and C. V. Aravind, "Device free human gesture recognition using Wi-Fi CSI: A survey," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103281.

[21] X. Wang, R. Huang, C. Yang, and S. Mao, "Smartphone sonar-based contact-free respiration rate monitoring," *ACM Trans. Comput. Healthcare*, vol. 2, no. 2, pp. 1–26, Apr. 2021.

[22] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 1–27, Sep. 2017.

[23] C. Lin et al., "A contactless authentication system based on WiFi CSI," *ACM Trans. Sensor Netw.*, vol. 19, no. 2, pp. 1–20, May 2023.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

[24] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[25] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[26] M. Li et al., "When csi meets public WiFi: Inferring your mobile phone password via WiFi signals," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 1068–1079.

[27] L. Lu et al., "An imperceptible eavesdropping attack on WiFi sensing systems," *IEEE/ACM Trans. Netw.*, vol. 32, no. 5, pp. 4009–4024, Oct. 2024.

[28] 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.

[29] 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, pp. 1–25, Jul. 2022.

[30] 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)*, Sep. 2022, pp. 10–18.

[31] X. Meng, J. Zhou, X. Liu, X. Tong, W. Qu, and J. Wang, "Secur-fi: A secure wireless sensing system based on commercial Wi-Fi devices," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, May 2023, pp. 1–10.

[32] J. Shenoy, Z. Liu, B. Tao, Z. Kabelac, and D. Vasisht, "RF-protect: Privacy against device-free human tracking," in *Proc. ACM SIGCOMM Conf.*, Aug. 2022, pp. 588–600.

[33] R. Ayyalasomayajula, A. Arun, W. Sun, and D. Bharadia, "Users are closer than they appear: Protecting user location from WiFi APs," in *Proc. 24th Int. Workshop Mobile Comput. Syst. Appl.*, Feb. 2023, pp. 124–130.

[34] Y. Yan and Z. Yang, "WiShield: Fine-grained countermeasure against malicious Wi-Fi sensing in smart home," in *Proc. Annu. Comput. Secur. Appl. Conf. (ACSAC)*, Dec. 2024, pp. 593–606.

[35] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, *arXiv:2010.02502*.

[36] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[37] X. Wang et al., "Placement matters: Understanding the effects of device placement for WiFi sensing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 1, pp. 1–25, 2022.

[38] Y. Liu, C. Du, T. Pang, C. Li, M. Lin, and W. Chen, "Graph diffusion policy optimization," 2024, *arXiv:2402.16302*.

[39] X. Xiang, S. Duan, H. Pan, P. Han, J. Cao, and C. Liu, "From one-hot encoding to privacy-preserving synthetic electronic health records embedding," in *Proc. Int. Conf. Cyberspace Innov. Adv. Technol.*, Dec. 2020, pp. 407–413.

[40] K. Black, M. Janner, Y. Du, I. Kostrikov, and S. Levine, "Training diffusion models with reinforcement learning," 2023, *arXiv:2305.13301*.

[41] A. Agarwal, S. M. Kakade, J. D. Lee, and G. Mahajan, "On the theory of policy gradient methods: Optimality, approximation, and distribution shift," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 4431–4506, Aug. 2019.

[42] V. Prakash Dwivedi and X. Bresson, "A generalization of transformer networks to graphs," 2020, *arXiv:2012.09699*.

[43] C. Vignac, I. Krawczuk, A. Siraudin, B. Wang, V. Cevher, and P. Frossard, "DiGress: Discrete denoising diffusion for graph generation," in *Proc. 11th Int. Conf. Learn. Represent.*, Jan. 2023, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=UaADNu86WX

[44] Y. Hu, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "DeFall: Environment-independent passive fall detection using WiFi," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8515–8530, Jun. 2022.

[45] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[46] Y. Zhang, Q. Liu, Y. Wang, and G. Yu, "CSI-based location-independent human activity recognition using feature fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[47] J. Huang et al., "Towards anti-interference WiFi-based activity recognition system using interference-independent phase component," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Jul. 2020, pp. 576–585.

[48] X. Lu, Y. Li, W. Cui, and H. Wang, "CeHAR: CSI-based channel-exchanging human activity recognition," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5953–5961, Apr. 2023.

[49] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[50] K. Sohn, X. Yan, and H. Lee, "Learning structured output representation using deep conditional generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Dec. 2015, pp. 3483–3491.

Jiacheng Wang received the Ph.D. degree from the School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. He is a Research Fellow with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests include wireless sensing, generative AI, semantic communications, and low-altitude wireless networks.

Hongyang Du (Member, IEEE) received the B.Eng. degree from Beijing Jiaotong University, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is an Assistant Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he directs the Network Intelligence and Computing Ecosystem (NICE) Laboratory. His research interests include edge intelligence, generative AI, and network management. He was a recipient of the IEEE ComSoc Young Professional Award for Best Early Career Researcher in 2024, the IEEE Daniel E. Noble Fellowship Award from the IEEE Vehicular Technology Society in 2022, the IEEE Signal Processing Society Scholarship from the IEEE Signal Processing Society in 2023, the Singapore Data Science Consortium (SDSC) Dissertation Research Fellowship in 2023, and the NTU Graduate College's Research Excellence Award in 2024. He was recognized as an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS. He serves as the Editor-in-Chief Assistant (2022–2024) and an Editor (since 2025) for IEEE COMMUNICATIONS SURVEYS AND TUTORIALS; an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY; and the Guest Editor for *IEEE Vehicular Technology Magazine*.

Yinqiu Liu received the B.Eng. degree from Nanjing University of Posts and Telecommunications, China, in 2020, and the M.Sc. degree from the University of California at Los Angeles, USA, in 2022. He is currently pursuing the Ph.D. degree with the College of Computing and Data Science, Nanyang Technological University, Singapore. His current research interests include blockchain security, mobile AIGC, and generative AI.

Geng Sun (Senior Member, IEEE) received the B.S. degree in communication engineering from Dalian Polytechnic University in 2011 and the Ph.D. degree in computer science and technology from Jilin University in 2018. He was a Visiting Researcher with the School of Electrical and Computer Engineering, Georgia Institute of Technology, USA. He is a Professor with the College of Computer Science and Technology, Jilin University. His research interests include wireless networks, UAV communications, mobile edge computing, and generative AI.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.

Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is a Professor with the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests are in the areas of sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

Shiwen Mao (Fellow, IEEE) is a Professor and the Earle C. Williams Eminent Scholar Chair, and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grids. He received the IEEE ComSoc MMTC Outstanding Researcher Award in 2023, the IEEE ComSoc TCCSR Distinguished Technical Achievement Award in 2019, and the NSF CAREER Award in 2010. He was a co-recipient of the 2022 Best Journal Paper Award of IEEE ComSoc eHealth Technical Committee, the 2021 Best Paper Award of Elsevier/Digital Communications and Networks (KeAi), the 2021 IEEE Internet of Things Journal Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the 2018 ComSoc MMTC Best Journal Paper Award, the 2017 Best Conference Paper Award, the 2004 IEEE Communications Society Leonard G. Abraham Prize in the field of communications systems, and several ComSoc technical committee and conference best paper/demo awards. He is the Editor-in-Chief of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He is a Distinguished Lecturer of IEEE Communications Society and the IEEE Council of RFID.

Dong In Kim (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California at Los Angeles, CA, USA, in 1990. He was a tenured Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. He is currently a Distinguished Professor with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. He is a fellow of Korean Academy of Science and Technology and a Life Member of the National Academy of Engineering of Korea. He was the first recipient of the NRF of Korea Engineering Research Center in Wireless Communications for RF Energy Harvesting from 2014 to 2021. He received several research awards, including the 2023 IEEE ComSoc Best Survey Paper Award and the 2022 IEEE Best Land Transportation Paper Award. He was selected the 2019 recipient of the IEEE ComSoc Joseph LoCicero Award for Exemplary Service to Publications. He was the General Chair of the IEEE ICC 2022, Seoul. He has been listed as a 2020/2022 Highly Cited Researcher by Clarivate Analytics. From 2001 to 2024, he served as an Editor, an Editor-at-Large, and an Area Editor of Wireless Communications I for IEEE TRANSACTIONS ON COMMUNICATIONS. From 2002 to 2011, he served as an Editor and a Founding Area Editor of Cross-Layer Design and Optimization for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. From 2008 to 2011, he served as the Co-Editor-in-Chief for the IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS. He served as the Founding Editor-in-Chief for IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2015.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include network resource management, wireless network security, the Internet of Things, 5G and beyond, and vehicular ad hoc and sensor networks. He is a registered Professional Engineer of Ontario, Canada; a fellow of the Engineering Institute of Canada, Canadian Academy of Engineering, and the Royal Society of Canada; a Foreign Member of Chinese Academy of Engineering; and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society. He was a recipient of the Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021; the R. A. Fessenden Award from IEEE, Canada, in 2019; the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019; the James Evans Avant Garde Award from the IEEE Vehicular Technology Society in 2018; the Joseph LoCicero Award and Education Award from the IEEE Communications Society in 2015 and 2017, respectively; and the Technical Recognition Award from Wireless Communications Technical Committee in 2019 and the AHSN Technical Committee in 2013. He was also a recipient of the Excellent Graduate Supervision Award from the University of Waterloo in 2006; and the Premier's Research Excellence Award (PREA) from the Province of Ontario, Canada, in 2003. He was the Technical Program Committee Chair/Co-Chair of IEEE Globecom 2016, IEEE Infocom 2014, IEEE VTC 2010 Fall, and IEEE Globecom 2007; and the Chair of IEEE Communications Society Technical Committee on Wireless Communications. He is the President of the IEEE Communications Society. He was the Vice President of Technical and Educational Activities, the Vice President of Publications, the Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a member of IEEE Fellow Selection Committee of the ComSoc. He was the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, *IEEE Network*, and *IET Communications*.

Authorized licensed use limited to: University of Waterloo. Downloaded on June 16,2025 at 15:27:09 UTC from IEEE Xplore. Restrictions apply.