Received 3 December 2024; accepted 29 December 2024. Date of publication 31 December 2024; date of current version 28 January 2025.

Digital Object Identifier 10.1109/OJCOMS.2024.3524497

Concept Drift Aware Wireless Key Generation in Dynamic LiFi Networks

ELMAHEDI MAHALAL [1] (Member, IEEE), ESLAM HASAN [2] (Graduate Student Member, IEEE), MUHAMMAD ISMAIL [2,3] (Senior Member, IEEE), ZI-YANG WU [4] (Member, IEEE), MOSTAFA M. FOUDA [5,6] (Senior Member, IEEE), ZUBAIR MD FADLULLAH [7] (Senior Member, IEEE), AND NEI KATO [8] (Fellow, IEEE)

[1]Department of Electrical, Computer Engineering and Computer Science, University of New Haven, West Haven, CT 06516, USA
[2]Department of Computer Science, Tennessee Tech University, Cookeville, TN 38505, USA
[3]Cybersecurity Education, Research, and Outreach Center, Cookeville, TN 38505, USA
[4]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[5]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA
[6]Center for Advanced Energy Studies, Idaho Falls, ID 83401, USA
[7]Department of Computer Science, Western University, London, ON N6A 5B7, Canada
[8]Graduate School of Information Sciences, Tohoku University, Sendai 980-8577, Japan

CORRESPONDING AUTHOR: E. MAHALAL (e-mail: emahalal@newhaven.edu)

This work was supported by NSF under Award 2138234.

ABSTRACT This paper studies the generation of cryptographic keys from wireless channels in light-fidelity (LiFi) networks. Unlike existing studies, we account for several practical considerations (a) realistic indoor multi-user mobility scenarios, (b) non-ideal channel reciprocity given the unique characteristics of the downlink visible light (VL) and uplink infrared (IR) channels, (c) different room occupancy levels, (d) different room layouts, and (e) different receivers' field-of-view (FoV). Since general channel models in dynamic LiFi networks are inaccurate, we propose a novel deep learning-based framework to generate secret keys with minimal key disagreement rate (KDR) and maximal key generation rate (KGR). However, we find that wireless channels in LiFi networks exhibit different statistical behaviors under various conditions, leading to concept drift in the deep learning model. As a result, key generation suffers from (a) a deterioration in KDR and KGR up to 29% and 38%, respectively, and (b) failing the NIST randomness test. To enable a concept drift aware framework, we propose an adaptive learning strategy using the similarity of channel probability density functions and the mix-of-experts ensemble method. Results show our adaptive learning strategy can achieve stable performance that passes the NIST randomness test and achieves 8% KDR and 89 bits/s KGR for a case of study with 60° FoV.

INDEX TERMS Concept drift, channel reciprocity, deep learning, infrared channel, key disagreement rate (KDR), key generation rate (KGR), light-fidelity (LiFi), ensemble strategy, multi user mobility, NIST randomness test, visible light communication (VLC), wireless secret key generation.

I. INTRODUCTION

WIRELESS data transmission over the open-air interface makes networks susceptible to passive attacks, such as eavesdropping, where attackers intercept ongoing communications to extract private information [1]. Common defense mechanisms involve encrypting data at the sender and decrypting at the receiver using symmetric/asymmetric key encryption, offering secure communications crucial for next-generation 5G+ networks [2]. Recently, wireless secret key generation (WSKG) has been explored, leveraging wireless channel randomness to generate secret keys without third-party intervention, thereby, enhancing security and efficiency [3], [4], [5].

There are several key advantages of WSKG over traditional cryptographic key generation methods, such as public key infrastructure (PKI) or symmetric key systems:

- Firstly, WSKG allows for the generation of long cryptographic keys with significantly less computational complexity. This is different from traditional methods that often require complex computations and substantial processing power for key generation and exchange. Also, WSKG leverages the randomness of the wireless channel, this natural variability provides a rich source of entropy that can be utilized efficiently to generate keys with minimal computational requirements [6].
- Secondly, WSKG offers the advantage of dynamically varying the key. This capability for frequent and automatic key updates enhances the security of the communication channel, making it more resilient to attacks compared to static keys used in traditional cryptographic methods.
- Finally, WSKG operates without the need for a third-party authority for key distribution, which is a fundamental requirement in traditional PKI systems. By eliminating the third-party intermediary, WSKG not only reduces potential points of vulnerability but also simplifies the system architecture, which leads to reductions in both operational costs and latency [7]. This decentralized aspect of WSKG is particularly valuable in scenarios where rapid deployment and independence from established infrastructure are beneficial. The traditional key process happens in the application layer, while in WSKG the extraction of the channel takes place in the physical layer, making it faster. In addition, traditional algorithms face significant limitations at the user equipment (UE) due to their high communication overhead and complexity. This complexity not only escalates device costs but also makes it impractical for low-cost terminals like the Internet-of-Things (IoT), Internet-of-Vehicles (IoV), Autonomous Aerial Vehicles (AAVs), and massive machine-type communication systems [8], [9], [10]. These systems are inherently sensitive to delays, constrained by power, or limited in processing capabilities, making the use of traditional cryptographic algorithms infeasible.

Given that 80% of data traffic originates indoors [11], our study focuses on indoor WSKG. Unlike the majority of existing research, this paper studies WSKG in a challenging 5G+ network, namely, the light-fidelity (LiFi) network. LiFi networks employ visible light communication (VLC) in the downlink (DL) and Infrared (IR) communication in the uplink (UL), hence, exhibiting non-ideal channel reciprocity given these correlated but distinct channels [12]. This challenge calls for optimization methods to minimize the key disagreement rate (KDR) between the uplink and downlink channels so that the key generation rate (KGR) can be maximized after information reconciliation is applied. Yet, a general LiFi channel model does not exist under user mobility [13]. This is because the characteristics of the blockage-sensitive LiFi channels are dominated by the dynamic interactions of the UE with the environment and other blockers (e.g., other users and furniture pieces in an indoor setup). Consequently, data-driven deep learning-based approaches should be adopted instead of model-based methods to optimize the WSKG framework. However, relevant works overlook the fact that deep learning methods are challenged by their generalization ability under practical considerations that include different room occupancy levels, room layouts, and receivers' field-of-view (FoV). To close this gap, our paper addresses these challenges and proposes an adaptive WSKG framework in practical LiFi networks.

A. RELATED WORK

The majority of existing research focuses on WSKG in radio channels [14], [15], [16] or mmWave channels [8], [9], [17], [18], [19], [20], [21], [22]. A few works have explored WSKG in the optical band, e.g., [23] and [24]. Most of the prior research assumes perfect channel reciprocity [17], [25], [26], [27], [28]. However, some recent works investigated scenarios of non-ideal channel reciprocity in both radio [29], [30] and mmWave [18] channels. Several existing studies have employed deep machine learning (ML) techniques in WSKG. For instance, the work in [31] introduced a WSKG framework based on deep learning for time division duplex systems and considered asynchronous channel state information measurements and hardware discrepancies. Also, the framework proposed in [31] utilized a feature extraction network consisting of two jointly trained auto-encoders, resulting in improved KDR compared to existing methods. Additionally, the research in [32] introduced a WSKG framework that relies on randomized pilots and long-short-term-memory (LSTM) recurrent neural networks (RNNs). This approach aims to enhance randomness distillation and mitigate man-in-the-middle attacks. Furthermore, in [33], an ML-based adaptive quantization level prediction scheme was developed to optimize the KGR with an accuracy of 98.2%. Moreover, the study in [34] employed neural networks to extract implicit features of wireless channels to generate secret keys adaptive to hardware inaccuracies. In [35] and [36], WSKG schemes were introduced using deep learning for feature mapping across different frequency bands in frequency division duplexing systems, which are usually non-reciprocal. Also, a WSKG scheme is designed in [37] to secure deep learning-based semantic communications by transmitting only the meaning of data instead of the raw message. In [38], an adversarial autoencoder is introduced for WSKG while preventing key leakage. In our previous work [39], we explored foundational techniques related to this study, which have been further developed and expanded in the current paper.

Limitations: To the best of our knowledge, only [23] has explored WSKG in dynamic LiFi networks that account for indoor user mobility. Without optimizing the KDR to address channel non-ideal reciprocity, the results in [23] showed a KDR of 40% and a KGR of 5 bits/s, indicating the need for significant improvement. In addition, while some existing studies employ deep learning techniques, they

focus on lower frequencies (2.4 and 2.5 GHz) in static environments. Hence, existing works do not account for the impact of user mobility on the performance of WSKG, which is crucial in 5G+, especially at higher frequencies. In such high-frequency bands, user mobility introduces blockage events, leading to channel outages and affecting the distribution of the channel impulse response (CIR). As a result, the effects of user mobility under different user densities/room occupancy levels, various FoV configurations, and different room layouts on deep learning-based WSKG in high-frequency bands are not studied in the literature. Furthermore, no countermeasures are studied to address possible negative effects.

B. CONTRIBUTION

The contributions of this paper involve (a) analysis of the impact of dynamic environments on deep learning-based WSKG models, highlighting performance deterioration, which is attributed to concept drift as will be explained in this paper, and (b) proposal of a robust deep learning-based WSKG model motivated by the analysis, showcasing its superior performance in dynamic environments. To elaborate, we carried out the following contributions to address the aforementioned research gaps in 5G+ WSKG:

- We propose a deep learning-based framework for WSKG in *indoor LiFi networks* with *multi-user mobility*. The framework incorporates a realistic model that emulates indoor human mobility, capturing macro and micro-mobility patterns. Mobility traces are used while generating CIR data, hence, considering link outages due to mobility-related events, i.e., link blockage and transmitter-receiver misorientation. A deep LSTM model at the access point (AP) is then trained to minimize the KDR of the preliminary keys by predicting quantized CIR levels that align closely with those at the UE side, thus, maximizing the KGR after information reconciliation.
- We assess the impact of *multi-user mobility*, *various user density/room occupancy levels*, *different FoVs*, and *different room layouts* on WSKG in indoor LiFi networks, using three performance metrics: KDR, KGR, and the ability to pass the NIST randomness tests. Experimental results show that our proposed LSTM model trained using CIR data from a fixed user density, FoV, and room layout achieves a KDR of 7% (33% improvement over [23]), average KGR of 89 bits/s, and the generated keys pass the NIST randomness tests. However, we observe that as the density of the users, the FoVs, and the layouts of the rooms change, the KDR deteriorates up to 29%, the KGR drops by 60% on average, and the generated keys fail the NIST tests. This decline is attributed to the concept drift effect, where a shift in the probability distribution of outage events and CIR occurs under changes in user density, FoV, and room layout.
- To enable a robust model, we investigated custom and general models for WSKG. The custom models did not generalize, and the general model did not achieve the best performances. Therefore, we propose an online strategy based on ensemble models. To reduce the complexity of the ensemble model (the number of custom models), we examine the probability density function (PDF) of the CIR at different user densities, FoVs, and room layouts. Then, we cluster the data according to PDF similarity using the Kolmogorov–Smirnov (KS) test, which reduces the complexity of the ensemble model by 50%. After clustering, we studied different ensemble strategies and found out that the mix-of-experts (MoE) model maintains a stable performance where the KDR is kept at 7%, the KGR at 89%, and the generated keys pass the NIST tests with FoV 60°.

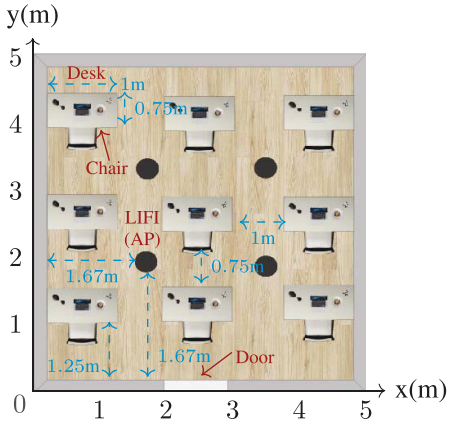
The rest of this paper is structured as follows. Section II introduces the system model. Section III outlines the proposed deep learning-based WSKG framework and presents the effect of dynamic environments on the deep learning model, showcasing its performance under varying conditions. This sets the stage for understanding the challenges faced by such models in dynamic environments. Section IV analyzes the reasons behind the behavior of deep learning models in dynamic settings and proposes an effective mitigation solution. Specifically, it presents our proposed online ensemble learning strategy based on PDF similarity and MoE model, and shows performance results. Finally, Section V concludes the paper and presents our future work.

II. SYSTEM MODEL

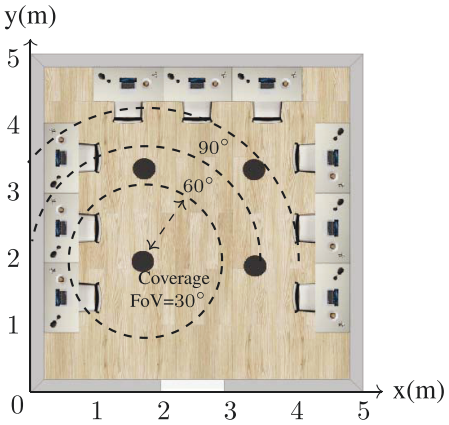
In this section, we provide an overview of the indoor setup, human mobility model, and channel model.

A. INDOOR SETUP

To study the generalization ability of the deep learning-based WSKG framework, we consider two layouts of an office room, denoted as R1 and R2. The room's dimensions are 5m×5m×3m. In R1, there are nine desks, with dimensions 1m×0.75m×1.3m, arranged as depicted in Figure 1(a). In R2, the same desks are distributed along the walls' sides, as shown in Figure 1(b). The room is covered by four LiFi APs evenly distributed across the ceiling, as illustrated in both layouts in Figure 1. We consider three distinct scenarios, each with a unique FoV for the UE, namely, 30°, 60°, and 90°. The coverage area of the APs is affected by these different FoVs, as illustrated in Figure 1(b). The human body is represented as a cuboid with dimensions of 1.8m×0.2m×0.45m. We consider a mass of 70 Kg for the human body, with a peak walking speed of 2.1 m/sec and a maximum acceleration of 1m/sec^2. The indoor mobility sample interval is set to 100 milli-sec. In our model, all surfaces in the room, as well as the human body, are considered reflectors and blockers of the line-of-sight (LoS) channels in both the uplink and downlink directions.



(a) Illustration of R1 layout



(b) Illustration of R2 layout

FIGURE 1. Office room setup showing the distribution of desks and LiFi APs. (a) showcases R1 with nine desks evenly distributed across the room, while (b) displays R2 with the desks placed along the three walls.

B. MOBILITY MODEL

We employ the indoor human mobility model detailed in [13], which accurately captures realistic human movements within indoor environments across two timescales: macro-scale and micro-scale. This model was validated against real indoor human mobility traces in [13], ensuring its reliability in practical applications. Relevant to this work, the adopted model is particularly valuable because it separates user trace data from channel data in the generation process, providing greater flexibility in simulating diverse mobility patterns.

1) MACRO-SCALE MOBILITY

Macro scale mobility patterns are captured using a semi-Markov process, which simulates movements among L destination points such as furniture and entrances within a space. These movements are driven by the return regularity and truncated Lévy walks, which dictate the likelihood of moving from one point to another based on historical visits and duration spent at each point. The process is mathematically represented as $\{(X_n, T_n) : n \geq 0\}$, where X_n denotes the location at the n-th transition and T_n signifies the time at the n-th transition. The semi-Markov kernel, detailing the transition probability from location i to j within time t, is expressed as:

$$\begin{aligned}\zeta_{i,j}^{(T_d)}(t) &= p_{i,j}^{(T_d)} R_{i,j}^{(T_d)}(t) \\ &= \Pr\{X_{n+1} = j, \Delta T_n \leq t | X_n = i, \Delta T_n \in T_d\}, \quad (1)\end{aligned}$$

where $R_{i,j}^{(T_d)}(t)$ models the time residence at location i before transitioning to j and $\Delta T_n = T_{n+1} - T_n$. The transition probabilities, $p_{i,j}^{(T_d)}$, vary depending on the time of day T_d (e.g., morning, afternoon).

Statistical results in the literature collectively indicate that human mobility follows a scale-free pattern, where human trajectories, regardless of the scale, display behaviors similar to those observed in a Lévy walk [40], [41]. The Lévy-walk component, defined by a truncated Pareto distribution, explains movements without returns, highlighting the long-tail characteristic of human movement. The PDF for transitioning from i to j, encompassing both step length and duration, is given by:

$$\Phi\left(r_{i\to j}, t_{i\to j}\right) = \phi\left(t_{i\to j} | r_{i\to j}\right) p\left(r_{i\to j}\right), \quad (2)$$

where $\phi(t_{i\to j}|r_{i\to j})$ is a conditional probability that a step takes $t_{i\to j}$ time in movement, and the step length probability follows a truncated Pareto distribution:

$$p_{i,j} = p\left(r_{i\to j}\right) = \frac{\alpha (r_{\min})^{\alpha} r_{i\to j}^{-(\alpha+1)}}{1 - (r_{\min}/r_{\max})^{\alpha}}, \quad (3)$$

with α as a positive parameter and $r_{\min}$ and $r_{\max}$ representing the minimum and maximum step lengths, respectively. The duration of stay at each location adheres to a similar truncated Pareto distribution, capturing the essence of human mobility on a macro scale.

2) MICRO-SCALE MOBILITY

The micro-scale mobility patterns are captured by the function $\Theta(i, j, t)$, which details the trajectories from starting points to destinations within a two-level framework. This framework consists of large-scale pathways constructed through sequences of next nodes and small-scale that detail the three-dimensional positioning and orientations of users.

For scenarios involving mobile states with mobility-impacted link quality, large-scale movements are found by calculating the shortest path using Dijkstra's algorithm. This is employed by two graphs: G_r for obstacles like furniture and G_p for navigable path nodes and avoiding no-go spots, thereby ensuring realistic movement trajectories. The shortest path between two locations i and j is determined using Dijkstra's algorithm $\mathcal{D}$ as follows:

$$\mathcal{V}_{i\to j} = \mathcal{D}(i, j), \quad (4)$$

To incorporate small-scale mobility, steering behaviors are utilized to model user-environment interactions, simulating authentic physical movements along the computed path nodes in an environment Ω:

$$\Theta(i, j, t) \ = \mathcal{S}\left(\mathcal{V}_{i\to j}, \Omega\right). \quad (5)$$

Users are represented as point masses with defined mass m, maximum acceleration $a_{\max}$, and maximum velocity $v_{\max}$, adhering to Newton's Second Law of Motion. The steering forces applied at time t result in acceleration $a(t) = \frac{F(t)}{m}$, with velocities and positions updated via Euler integration:

$$v(t) = v(t - \delta\tau) + a(t)\delta\tau, \tag{6}$$

$$p(t) = p(t - \delta\tau) + v(t)\delta\tau. \tag{7}$$

This framework incorporates seek and avoidance behaviors, guiding the user towards targets while avoiding obstacles. The seek behavior produces a seek force, $\mathbf{F}_s(t)$, which attracts the user to each target node. The seek force orients towards the distance vector $\mathbf{d}(t)$ between the intermediate target $\xi(t) \in \mathcal{V}_p$ and the actual position of the user $\mathbf{p}(t)$ as $\mathbf{d}(t) = \xi(t) - \mathbf{p}(t)$. The corresponding desired velocity vector $\mathbf{v}_d(t)$ is given by

$$\mathbf{v}_d(t) = \frac{\mathbf{d}(t)}{\|\mathbf{d}(t)\|} v_{\max}\delta\tau. \tag{8}$$

The user is driven by the seeking force $\mathbf{F}_s(t)$ as

$$\mathbf{F}_s(t) = m\left(\frac{\mathbf{v}_d(t) - \mathbf{v}(t)}{\delta\tau}\right). \tag{9}$$

As the user approaches the destination, they slow down to end the period of mobility. This behavior is modeled by an arrival force $\mathbf{F}_a(t)$ that is opposite to $\mathbf{F}_s(t)$ but has a threshold radius to ignore long-range effects. To repulse a user from penetrating insurmountable areas such as obstacles, an avoidance force $\mathbf{F}_o(t)$ is applied. The avoidance force is calculated considering the perpendicular distance d_w from the present position to the surface of an obstacle. This is given by

$$\mathbf{F}_o(t) = m\left(\frac{\mathbf{n}_w d_w \|\mathbf{v}(t) - \mathbf{v}(t - \delta\tau)\|}{\delta\tau}\right), \tag{10}$$

where $\mathbf{n}_w$ denotes an orthogonal unit vector against the obstacle surface w. The overall force applied to the user at any time t combines these behaviors:

$$\mathbf{F}(t) = \mathbf{F}_s(t) + \gamma_o \sum_{w}^{\mathcal{W}} \mathbf{F}_o(t) + \gamma_a \mathbf{F}_a(t). \tag{11}$$

Here, γ_o and γ_a are tuning parameters that adjust the influence of the avoidance and arrival efforts, respectively. This integrated approach ensures that the user moves smoothly towards the target while avoiding obstacles and decelerating appropriately upon arrival.

The model further integrates the stochastic orientation of UEs, described by yaw, pitch, and roll, significantly affecting the optical wireless channel. Given its slow variation, which is attributed to minor delay spreads and extended coherence times, the orientation angles impact channel characteristics. Specifically, the polar angle ϑ adheres to a Laplace distribution for seated scenarios and a Gaussian distribution for ambulatory activities, influencing the azimuth angle ω following user paths [42].

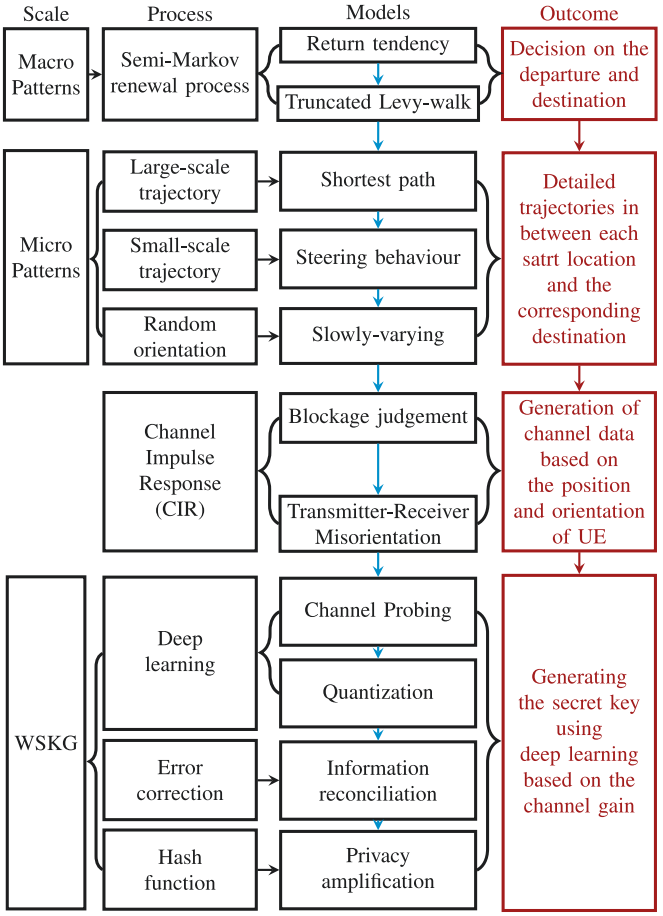


FIGURE 2. Framework for deep learning based WSKG in dynamic multi-user indoor LiFi.

This mobility modeling approach is depicted in the upper section of Figure 2. Further elaboration on the mobility model's parameters and the underlying algorithms is provided in [13].

C. CHANNEL MODEL

The LoS impulse response is expressed as follows [43], [44]:

$$h^{(0)}(t,\tau) = \begin{cases} \frac{A_R}{d_0^2(t)} \frac{(m+1)}{2\pi} \cos^m \psi(t) \cos\theta(t) \\ \times T_S(\theta(t)) \delta\left(\tau - \frac{d_0(t)}{c}\right), \\ \qquad \text{if } 0 \le \theta(t) \le \Psi \\ 0, \quad \text{if } \theta(t) > \Psi \text{ or ray is blocked,} \end{cases} \tag{12}$$

where τ denotes delay, A_R is the sensor area, ψ the angle of irradiance, Ψ the receiver's FoV, c the speed of light, and $T_S(\theta)$ the overall transmission response of the optical system (assumed to be 1 in this analysis). LoS transmission distance $d_0(t) = \|p(t) - p_{\text{AP}}(v)\|$, where $p_{\text{AP}}(v)$ denotes the AP's location, with v ranging over the set of all APs V. Additionally, the mode number m relates to the half-power angle $\Phi_{1/2}$ through $m = -\ln 2 / \ln \cos \Phi_{1/2}$. The receiver incidence angle θ is determined by ϑ and ω, and the relative angle between transmitter and receiver, considering the transceiver's placement on the terminal.

TABLE 1. Time and space complexities of WSKG tasks.

WSKG Task	Time Complexity	Space Complexity
Channel probing	$O(N)$	$O(N)$
Quantization	$O(2^Q + N2^Q)$	$O(N + 2^Q)$
Information reconciliation	$O(N)$	$O(N)$
Privacy amplification	$O(N)$	$O(1)$

To mitigate potential eye irradiation hazards from the uplink and blockages by the user's body, leading to significant outages, the UE's transceiver is oriented forward along the trace direction, not upward towards the ceiling. Thus, the terminal transceiver's direction vector is defined as $u_\theta^{\mathrm{UL}}(t) = (\sin\vartheta(t)\cos\omega(t), \sin\vartheta(t)\sin\omega(t), \cos\vartheta(t))$, and for an AP transceiver as $u_\theta^{\mathrm{DL}} = (0, 0, -1)$. The cosine of $\theta(t)$ is calculated as $\cos\theta(t) = u_\theta^{\mathrm{UL}}(t) \cdot u_\theta^{\mathrm{DL}} / (\|u_\theta^{\mathrm{UL}}(t)\| \|u_\theta^{\mathrm{DL}}\|)$. A ray is considered blocked if it intersects any surface, including furniture or the user's body. The channel model in (12) applies to both VL and IR and the reciprocity between uplink and downlink channels is influenced by environmental conditions and channel dynamics [13]. IR and VL bands differ primarily in their wavelengths. IR light has longer wavelengths than VL, which affects how each band interacts with physical environments, including absorption and reflection properties, as mentioned in [13, Table 1]. The reciprocity between uplink and downlink channels in these bands is influenced by several factors, including environmental conditions and channel dynamics.

III. PROPOSED DEEP LEARNING-BASED KEY GENERATION STRATEGY

In LiFi networks, the uplink and downlink channels operate in different frequency bands, namely the IR and VL bands, respectively. Since these channels exhibit non-ideal reciprocal characteristics, the quantization thresholds at both the AP and UE must be optimized to generate closely matching preliminary keys after quantization. Failure to do so would result in a high KDR and a low KGR. However, accurate channel models do not exist in LiFi networks as the CIR is tied to the user's mobility details and interactions with the environment, making data-driven optimization of quantization thresholds essential. Hence, we propose a deep learning-based quantization strategy at the AP. Specifically, we train a classification LSTM model at the AP using a dataset of normalized uplink CIR $\bar{h}_{\mathrm{UL}}(t)$ as input features. The model output corresponds to the relevant downlink Gray code generated with Algorithm 1 (to be detailed next). The model is trained to minimize categorical cross-entropy, effectively, reducing the KDR. This means that the model learns to relate $\mathbf{H}_{\mathrm{UL}}$ (used at the AP to generate the downlink preliminary key $\tilde{K}_{\mathrm{DL}}$) with $\tilde{K}_{\mathrm{UL}}$ (preliminary key generated at the UE). In testing, the model is provided with the uplink CIR from the UE, and it generates a preliminary key at the AP that closely matches (i.e., minimizes KDR) the UE's preliminary key. Figure 3 illustrates the proposed WSKG approach. The details are provided next.

Algorithm 1 CDF-Based Quantization

Input: $\mathbf{H}$ %Estimate of CIR
Input: Q %Quantization level
Output: $\tilde{K}$ %Generated preliminary key sequence
procedure CDF_QUANTIZATION($\mathbf{H}$, Q)
  $F(h) \leftarrow \Pr(\mathbf{H} < h)$ %CDF calculation
  $\eta_0 \leftarrow -\infty$ %Threshold
  for $j \leftarrow 1$ to $2^Q - 1$ **do**
    $\eta_j \leftarrow F^{-1}(j/2^Q)$ %Threshold
  end for
  $\eta_{2^Q} \leftarrow \infty$
  Construct Gray codes b_j and assign them to different intervals $[\eta_{j-1}, \eta_j]$
  for $n \leftarrow 1$ to N **do**
    if $\eta_{j-1} \le h(n) < \eta_j$ **then**
      $\tilde{K}(t, Q) \leftarrow b_j$
    end if
  end for
end procedure

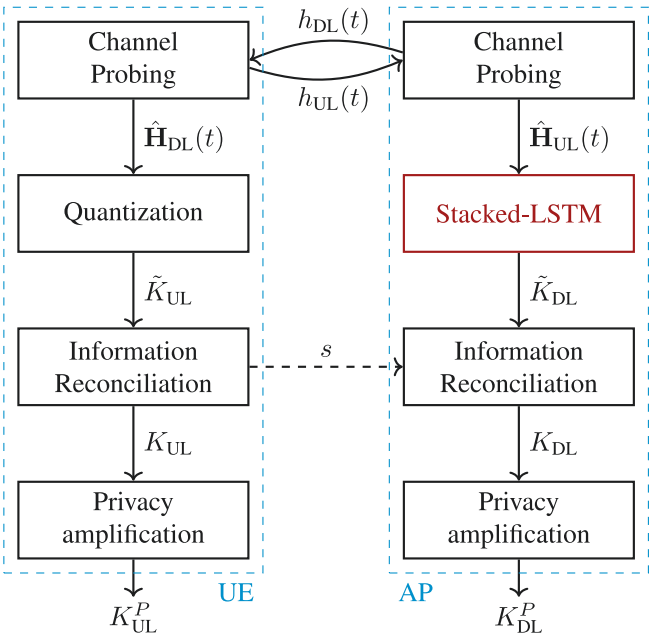


FIGURE 3. Further detailed illustration of the key generation algorithm depicted in the bottom part of Figure 2.

A. WSKG PROCEDURE

The bottom part of Figure 2, detailed in Figure 3, illustrates the proposed framework for WSKG in mobile LiFi networks. The process begins by generating indoor human mobility traces, as explained in Section II-B. Subsequently, we generate the CIR data. This is done by first performing blockage judgment to determine if the link between the user

and the AP is blocked. A ray is considered blocked if it intersects any of the six surfaces representing each cuboid object in the room (e.g., human or furniture). Additionally, we verify if the received signal falls within the receiver's FoV. If the ray is unblocked and the signal is within the FoV, the CIR is computed according to (12). Then, we use the CIR data to generate the secret key based on channel probing, quantization, and information reconciliation processes. We will discuss these concepts and then introduce our proposed deep learning approach (Figure 3) to improve the KDR and KGR in the uplink and downlink keys caused by channel non-ideal reciprocity.

1) CHANNEL PROBING

Initially, the user and AP engage in bidirectional communication by exchanging request and response probing frames over a specified duration. Upon receiving a request frame, the receiver provides a reply frame. The time interval between consecutive request or reply probing frames is assumed to be constant and denoted as Δ, resulting in a channel probing rate of $1/\Delta$. At the end of the channel probing process, the user and AP have collected a set of N pairs of channel measurements, in this work $N = 100$. The estimated channel gains in the downlink and uplink are given by

$$\begin{cases} \mathbf{H}_{\mathrm{DL}} = [h_{\mathrm{DL}}(1), h_{\mathrm{DL}}(2), \ldots, h_{\mathrm{DL}}(N)]^{\mathrm{T}}, \\ \mathbf{H}_{\mathrm{UL}} = [h_{\mathrm{UL}}(1), h_{\mathrm{UL}}(2), \ldots, h_{\mathrm{UL}}(N)]^{\mathrm{T}}, \end{cases} \tag{13}$$

where $h(n)$ represents the CIR estimate at discrete instances $1 \leq n \leq N$. The notation T denotes the transpose. It should be highlighted that a perfect estimation of CIR is assumed.

2) QUANTIZATION

We adopt a cumulative distribution function (CDF)-based quantization approach. This method determines quantization thresholds based on the CDF of the CIR data. The use of CDF-based quantization ensures an equal balance of "1"s and "0"s, which is crucial for passing the NIST randomness tests. Our decision to use the CDF-based quantizer stems from preliminary experiments that showed the distribution of CIR shifts with different user densities. Algorithm 1 presents a description of the CDF-based quantizer. The inputs to the algorithm include the CIR estimate $\mathbf{H}$ over the probing interval and the desired quantization level Q, we set $Q = 2$. The output is the preliminary key sequence $\tilde{K}$. The quantization process starts by calculating the thresholds (η_j for $0 \leq j \leq 2^Q$) based on the CDF of the CIR (lines $5-10$ in Algorithm 1). Subsequently, Gray codes b_j are assigned to each threshold interval $[\eta_{j-1}, \eta_j]$ (line 11). Finally, each CIR sample within a given quantization interval is replaced by the corresponding Gray code. This approach can be extended to multi-bit quantization by incorporating more quantization levels. To maintain a Hamming distance of one between similar data samples, a Gray code is employed, resulting in similar binary strings with only a one-bit difference.

Typically, the UE uses Algorithm 1 to quantize $\mathbf{H}_{\mathrm{DL}}$ (received from the AP) to create its preliminary key $\tilde{K}_{\mathrm{DL}}$.

Algorithm 2 Information Reconciliation

Input: $\tilde{K}_{\mathrm{DL}}$, $\tilde{K}_{\mathrm{UL}}$ %Quantized keys of AP and UE
Input: C %ECCA set shared by AP and UE
Output: K_{DL}, K_{UL} %Reconciled key
1: UE randomly selects a code c from the ECCA set C
2: UE calculates the syndrome $s = \mathrm{XOR}(\tilde{K}_{\mathrm{DL}}, c)$ and transmits s to the AP through a public channel
3: UE assigns $K_{\mathrm{DL}} = \tilde{K}_{\mathrm{DL}}$
4: AP receives s and calculates $\tilde{c}_{\mathrm{AP}} = \mathrm{XOR}(\tilde{K}_{\mathrm{UL}}, s)$
5: AP decodes $\tilde{c}_{\mathrm{AP}}$ to get c_{AP}
6: AP calculates $K_{\mathrm{UL}} = \mathrm{XOR}(c_{\mathrm{AP}}, s)$

Similarly, the AP employs Algorithm 1 to quantize $\mathbf{H}_{\mathrm{UL}}$ (received from the UE) and generates its preliminary key $\tilde{K}_{\mathrm{UL}}$. In this work, to generate the dataset needed to develop the proposed LSTM model, Algorithm 1 is used in the uplink and downlink. Once our proposed deep learning-based approach is developed using the created dataset, Algorithm 1 is adopted at the UE level, while the deep learning model is used at the AP level for key generation (see Figure 3).

3) INFORMATION RECONCILIATION

This step creates identical symmetric keys at the UE and AP to be used in cryptography. This is done by identifying and removing non-identical bits in the preliminary keys between the AP and UE, denoted as $\tilde{K}_{\mathrm{DL}}$ and $\tilde{K}_{\mathrm{UL}}$, to generate the final symmetric keys K_{DL} and K_{UL}. Our focus is on the KDR which describes the percentage of non-identical bits between $\tilde{K}_{\mathrm{DL}}$ and $\tilde{K}_{\mathrm{UL}}$ as these need to be removed. Our aim in this paper is to reduce the KDR due to channel non-ideal reciprocity between the uplink and downlink to enhance the efficiency of the key generation process and maintain a high KGR. The secure sketch is a widely used error correction code-based approach (ECCA) for information reconciliation [45], as described in Algorithm 2. The process starts with UE randomly selecting a codeword c from the Bose–Chaudhuri–Hocquenghem (BCH) code set C. A BCH (w, m, e) code consists of a w-bit codeword and an m-bit message, with the capability to correct up to e-bit errors. Then, the UE calculates the syndrome s using the exclusive-OR (XOR) operation, given by $s = \mathrm{XOR}(\tilde{K}_{\mathrm{DL}}, c)$. UE transmits the syndrome s to the AP. Then, the AP calculates a codeword $\tilde{c}_{\mathrm{AP}} = \mathrm{XOR}(\tilde{K}_{\mathrm{UL}}, s)$. If the errors are correctable, i.e., e-error bits or less, AP can decode $\tilde{c}_{\mathrm{AP}}$ to obtain c_{AP}, such that $c_{\mathrm{AP}} = c$. Finally, it derives a new key through the XOR operation, denoted as $K_{\mathrm{UL}} = \mathrm{XOR}(c_{\mathrm{AP}}, s)$.

4) PRIVACY AMPLIFICATION

During the process of probing, quantization, and information reconciliation, the UE and AP communicate over public channels, making them susceptible to potential eavesdropping by attackers. To address this security concern, a privacy amplification step is commonly incorporated after information reconciliation, where universal hash families are

used. In this paper, we use SHA-256, which is part of the SHA-2 family, designed by the National Security Agency (NSA) and published in 2001 by the National Institute of Standards and Technology (NIST) [46].

Finally, we provide the time and space complexity of our WSKG procedure in Table 1. The time complexity is dominated by $O(2^Q + N2^Q)$ and the space complexity is dominated by $O(N + 2^Q)$. The latency introduced by WSKG is competitive compared to the latency inherent in traditional cryptographic exchanges, which often involve multiple rounds of communication between endpoints and a central authority [47].

B. DATASET GENERATION

The framework presented in Figure 2 is utilized to generate the necessary uplink and downlink CIR datasets. All data generation processes were conducted on the Tennessee Technological University's High-Performance Computing (HPC) cluster. The generation of indoor mobility traces was based on the mobility model outlined in Section II-B. The CIR data is collected under various scenarios of user densities in rooms R1 and R2, namely, with $1, 3, 6$, and 8 mobile users in each room. For each scenario, three different FoVs of $30°, 60°$, and $90°$ are considered. To prevent overfitting, in each scenario, 1000 mobility traces are generated per user, and then the CIR data is collected as described above. As a result, time-series data of uplink and downlink CIR are recorded for each user and AP, as in (13).

It is essential to emphasize that the CIR undergoes three distinct stages: the *Entering Stage* when the user enters the room and moves to the first residence spot, the *Wandering Stage* when the user roams across the room, and the *Exiting Stage* when the user moves toward the exit. In this study, we specifically concentrate on the *Wandering Stage* since it is the extended phase that triggers the WSKG process.

Before training the LSTM model, data pre-processing is carried out to extract the necessary input features and labels.

C. DATASET PRE-PROCESSING

The following pre-processing steps are undertaken to define the input features and the output class.

1) NORMALIZING INPUT FEATURES

To ensure convergence during the model training, the uplink CIR data is normalized. This involves defining $h_{\text{UL}}^{\max} = \max \mathbf{H}_{\text{UL}}$ and $h_{\text{UL}}^{\min} = \min \mathbf{H}_{\text{UL}}$ and then calculating the input features $\mathbf{X} = [x(1), x(2), \ldots, x(N)]$ as

$$x(t) = \frac{h_{\text{UL}}(n) - h_{\text{UL}}^{\min}}{h_{\text{UL}}^{\max} - h_{\text{UL}}^{\min}}. \tag{14}$$

2) DEFINING OUTPUT LABELS/CLASSES

To minimize the KDR, we first determine the quantization (Gray code) outcome at the UE based on the downlink CIR $\mathbf{H}_{\text{DL}}$. This is achieved by utilizing the CDF_QUANTIZATION($\mathbf{H}_{\text{DL}}, Q$) procedure in Algorithm 1. Consequently, at each time instant n, we obtain the corresponding Gray code outcome $\tilde{K}_{\text{DL}}(n)$. Since the Gray code represents binary outcomes, we find its decimal equivalent and use it as the output label/class $y(n)$ for the model at time n. Altogether, there are 2^Q thresholds mapped to 2^Q different Gray codes, resulting in 2^Q classes/labels. These classes/labels, corresponding to the input features, are represented as $\mathbf{Y}$.

D. STACKED-LSTM MODEL TRAINING AND OPTIMIZATION

The proposed model is built upon the stacked-LSTM architecture, which is particularly suited for handling time-series data like ours. We opt for LSTM due to its ability to address the challenges of exploding and vanishing gradients. The model is trained in a supervised manner using $(\mathbf{X}, \mathbf{Y})$ examples from the dataset. To minimize the categorical cross-entropy, equivalent to reducing the KDR between uplink and downlink channels, we utilize the Backpropagation-through-time (BPTT) algorithm. Additionally, we perform a grid search to optimize the model's hyper-parameters, including the number of hidden layers, the number of LSTM cells per hidden layer, and the activation functions. Through experimentation, we find that a configuration of stacking two LSTM layers yields optimal results. The first LSTM layer consists of 56 hidden units and utilizes the tanh activation function, the input size is a vector of the wandering stage $(55, 1)$, and the output is a return sequence, while the second LSTM layer comprises 72 hidden units with the Sigmoid activation function, it takes the output of the previous LSTM and returns the sequence. The output dense layer contains 2^Q neurons, corresponding to the number of classes, and applies the Softmax activation function. Our model is trained by setting the learning rate to 0.001, employing the Adam optimizer, employing a batch size of 10, and conducting 1000 epochs to achieve convergence.

Several mitigation techniques have been adopted to avoid overfitting. First, we ensured that the generated user trace datasets were highly diverse, producing distinct CIR scenarios to prevent overly similar training instances. Specifically, we generated 1000 diverse mobility traces under various user densities and in various room configurations and FoV settings. Additionally, we utilized an $60{-}20{-}20\%$ data split, dedicating 60% of the data to training, 20% for validation, and 20% to testing, thereby enabling the model to be evaluated on unseen data to confirm that the model is not overfitting.

This model avoids the iterative steps typically needed in traditional quantization methods described in Algorithm 1. This reduces both time and space complexity from $O(2^Q + N2^Q)$ and $O(N + 2^Q)$, respectively, to $O(N)$, allowing the model to make real-time predictions efficiently. The parameters tuned in our model include the number of LSTM layers, hidden units per layer, learning rate, batch size, and number of epochs. Each of these parameters is

selected to balance convergence speed, model complexity, and robustness in dynamic Li-Fi environments. For example, increasing the number of LSTM layers and hidden units enhances the model's ability to capture complex temporal patterns in channel data, which reduces the KDR. However, this also raises computational costs and risks overfitting, especially in low-density scenarios.

E. PERFORMANCE EVALUATION METRICS

During the testing phase, the model takes the normalized uplink CIR as input features and generates the downlink preliminary keys $\tilde{K}_{\mathrm{DL}}$ as its output. Then, information reconciliation is applied to derive the final keys. The performance of the model is then assessed using the following metrics:

1) KDR

Due to the inherent non-ideal reciprocity of uplink and downlink channels in LiFi networks, there is a possibility of having mismatched bits following quantization (before information reconciliation). The KDR is a metric that quantifies the ratio of mismatched bits present in the preliminary keys $\tilde{K}_{\mathrm{DL}}$ (generated at the AP) and $\tilde{K}_{\mathrm{UL}}$ (generated at the UE). These keys are produced using Algorithm 1 for downlink and the LSTM model for uplink. Assuming the length of bits in the preliminary keys is denoted by $L = Q \times N$, where N represents the length of channel measurements and Q is the quantization level. The KDR can be mathematically expressed as

$$\mathrm{KDR} = \frac{\sum_{l=1}^{L} |\tilde{K}_{\mathrm{UL}}(l) - \tilde{K}_{\mathrm{DL}}(l)|}{L}. \tag{15}$$

2) KGR

Following the process of information reconciliation, a uniform key K is established between the AP and the UE, resulting in a KDR value of zero. This metric is quantified in terms of bits/second and signifies the speed at which a system can generate or update keys. Different cryptographic algorithms necessitate specific KGR. For instance, the AES algorithm demands a KGR of 0.1 bit/second [14], [48].

3) KEY RANDOMNESS

Evaluating the resilience of generated keys against unauthorized access is paramount. To assess the randomness of the produced keys, we employ the NIST randomness tests, detailed in [49]. The outcome of each test is represented by a p-value. A key is considered to exhibit substantial randomness and successfully passes the respective test if the p-value is equal to or greater than 0.01. The suite of tests includes the monobit, frequency within the block, longest run of ones, binary matrix rank, discrete Fourier transform, non-overlapping template matching, overlapping template matching, approximate entropy, cumulative sums, and random excursion tests. The monobit test scrutinizes the distribution balance between 0s and 1s in a sequence. Meanwhile, the frequency within a block test assesses the equilibrium of 0s and 1s within discrete M-bit blocks. The longest run of one test evaluates the lengthiest streak of consecutive 1s within a block. The binary matrix rank focuses on the rank of separate sub-matrices. The discrete Fourier transform (DFT) investigates the peaks' magnitudes in the sequence's DFT. The non-overlapping template matching test searches for specific non-periodic patterns. The approximate entropy test measures the regularity of fluctuations in a sequence. The cumulative sums test detects any significant deviations in the tally of 0s and 1s across the sequence. Lastly, the random excursion test monitors the count of cycles with exactly T-visits in a cumulative sum random walk. A key is considered successful and passes our criteria only if it passes all 15 tests; otherwise, it fails.

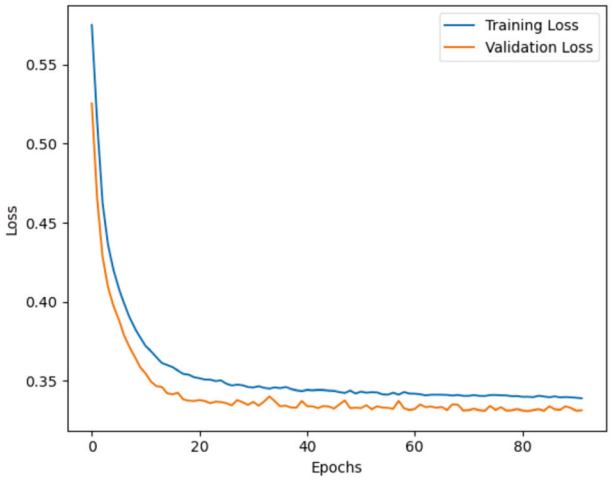


FIGURE 4. Learning curve for the DNN-based strategy.

F. PERFORMANCE EVALUATION RESULTS

This subsection provides an overview of the findings obtained through the performance evaluation. First, we motivate the adoption of LSTM-based model rather than the less complex deep feedforward neural network (DNN) model by comparing the learning curves of both models. Second, we test the generalization ability of the LSTM-based WSKG model. Finally, we compare a learning-based WSKG framework with a traditional benchmark.

1) LSTM VERSUS DNN MODEL

Herein, we compare two deep learning methods to motivate the use of an LSTM-based strategy. Specifically, we compare a DNN-based strategy for WSKG against an LSTM-based strategy, with the training and validation losses of both strategies summarized in Figure 4 and Figure 5, respectively. Each strategy adopts Algorithm 1 at the UE level, while the deep learning model (DNN or LSTM) is used at the AP level for WSKG. We performed hyper-parameter optimization for both LSTM and DNN models. The results demonstrate a steady decrease in the training and validation losses, showcasing no overfitting or underfitting. More importantly, our results indicate that the LSTM-based strategy exhibits

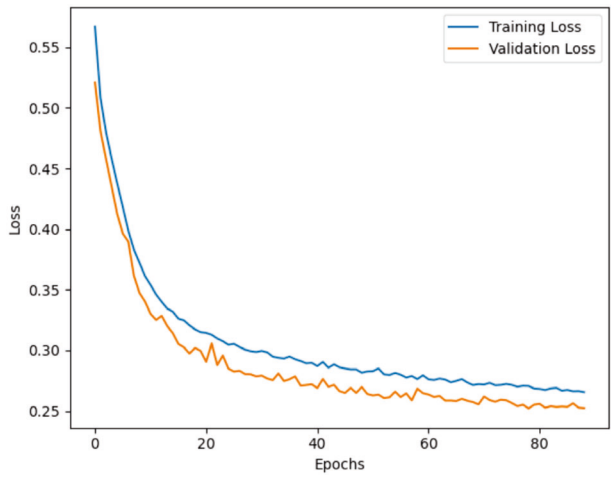


FIGURE 5. Learning curve for the LSTM-based strategy.

a loss performance that is 15% better than the DNN-based strategy. This performance is attributed to the nature of the LSTM model, which exploits the temporal correlation within the time-series data representing the CIR, which improves the loss performance compared with the DNN model.

2) GENERALIZATION ABILITY OF THE LSTM MODEL

To better understand the generalization ability of the proposed LSTM model, we trained and tested eight different models. The first six represent models that are trained using a specific density of users in the room within a specific FoV. Hence, models $\mathbf{M}_{1,30^\circ}$, $\mathbf{M}_{3,30^\circ}$, $\mathbf{M}_{6,30^\circ}$, $\mathbf{M}_{1,60^\circ}$, $\mathbf{M}_{3,60^\circ}$, and $\mathbf{M}_{6,60^\circ}$, are trained using $(\mathbf{X}, \mathbf{Y})$ samples collected from R1 when there are 1, 3, and 6 mobile users in the room with 30° and 60° FoV, respectively. Furthermore, we trained two other models $\mathbf{M}_{G,30^\circ}$ and $\mathbf{M}_{G,60^\circ}$ using $(\mathbf{X}, \mathbf{Y})$ samples from all the aforementioned mobile user densities in each FoV. In the performance evaluation, we test each model when there are 1, 3, and 6 mobile users in the room with two different FoVs 30° and 60°. The same procedure was done with R2 and for FoV 90° and the conclusions were consistent with what we will discuss next. Hence, for clarity of the presentation we summarize the KDR, KGR, and NIST randomness test results for the eight models we mentioned for R1 and FoVs 30° and 60° in Table 2.

The following remarks can be made based on Table 2:

- The bold diagonal elements in the table show the KDR, KGR, and NIST randomness tests when the model ($\mathbf{M}_{1,30^\circ}$, $\mathbf{M}_{3,30^\circ}$, $\mathbf{M}_{6,30^\circ}$, $\mathbf{M}_{1,60^\circ}$, $\mathbf{M}_{3,60^\circ}$, and $\mathbf{M}_{6,60^\circ}$) is trained and tested with the same number of users in the room (UE 1, 3, and 6). These diagonal elements gave the minimum KDR results ($14\% - 15\%$ and $7\% - 8\%$ in FoVs 30° and 60°, respectively) and maximum KGR ($78 - 80$ bits/s and $89 - 90$ bits/s in FoVs 30° and 60°, respectively). Also, these models passed all the NIST randomness tests.
- The off-diagonal elements indicate instances where models ($\mathbf{M}_{1,30^\circ}$, $\mathbf{M}_{3,30^\circ}$, $\mathbf{M}_{6,30^\circ}$, $\mathbf{M}_{1,60^\circ}$, $\mathbf{M}_{3,60^\circ}$, and $\mathbf{M}_{6,60^\circ}$) were trained and tested on different numbers of users. For example, the second cell in the first row shows the result when the model $\mathbf{M}_{1,30^\circ}$ was trained on the case where there is 1 UE in the room but tested on a case where there are 3 UE in the room. These results show deteriorating performance in KDR by 28% in FoV 30° and 24% in FoV 60°. Here, KGR drops are notable as well, by 38% for FoV 30° and 30% for FoV 60°. Also, these models fail some of the NIST randomness tests. For example, $\mathbf{M}_{1,30^\circ}$ fails the "overlapping template matching" test when there are 3 to 6 UEs. It also fails the "random excursion" test when there are 6 UEs.
- For the models termed $\mathbf{M}_G$, the KDR is between $39\% - 44\%$, and $23\% - 27\%$ in FoVs 30° and 60°, respectively, while the KGR is around 44 bits/s and 66 bits/s in FoVs 30° and 60°, respectively. This result is worse by around 20% than the models trained and tested on the same number of users. More importantly, the models $\mathbf{M}_G$ fail the following NIST randomness tests: "overlapping template matching" with 1 user, "block frequency" with 3 users, and both "random excursions" and "frequency within a block" with 6 users. Hence, the developed keys under them are not viable.
- When it comes to the dynamicity and mobility impact on deep learning-based WSKG, we observe that such dynamicity introduces dynamic channel blockages, which further disrupt the reciprocity between uplink and downlink channels. This dynamicity-induced non-reciprocity results in an increase in the KDR and a reduction in KGR.

TABLE 2. Summary of KDR, KGR and NIST results.

	KDR%			KGR(bits/s)			NIST		
Models	1UE	3UEs	6UEs	1UE	3UEs	6UEs	1UE	3UEs	6UEs
$\mathbf{M}_{1,30^\circ}$	**14.26**	41.97	42.79	**79.05**	42.23	41.46	**Passed**	Failed	Failed
$\mathbf{M}_{3,30^\circ}$	42.26	**15.18**	39.06	42.40	**77.84**	47.11	Failed	**Passed**	Failed
$\mathbf{M}_{6,30^\circ}$	43.92	38.46	**15.67**	40.36	47.49	**77.38**	Failed	Failed	**Passed**
$\mathbf{M}_{G,30^\circ}$	38.86	42.79	44.25	48.24	44.44	43.33	Failed	Failed	Failed
$\mathbf{M}_{1,60^\circ}$	**7.97**	28.07	31.12	**89.84**	63.00	59.53	**Passed**	Failed	Failed
$\mathbf{M}_{3,60^\circ}$	27.75	**8.43**	29.58	62.80	**88.90**	61.62	Failed	**Passed**	Failed
$\mathbf{M}_{6,60^\circ}$	29.12	27.38	**8.51**	61.02	64.23	**88.93**	Failed	Failed	**Passed**
$\mathbf{M}_{G,60^\circ}$	23.20	24.66	26.53	68.66	67.95	65.47	Failed	Failed	Failed

3) LSTM-BASED MODEL VERSUS TRADITIONAL BENCHMARK

Herein, we compare the performance of the proposed LSTM-based framework against a traditional method presented

TABLE 3. Comparison of KDR and KGR with and without deep learning.

Metric	Traditional Method of [23]	LSTM-based $\mathbf{M}_{1,30^\circ}$
KDR (%)	41.63	14.26
KGR (bps)	5.26	79.05

in [23] that adopts channel probing, quantization, and reconciliation without any deep learning-based performance optimization. The comparative study considers room R1 with FoV 30°. We selected the best-performing strategy from [23] and compared it under the same settings with our proposed model $\mathbf{M}_{1,30^\circ}$. Both methods pass the NIST randomness tests under stationary conditions (i.e., no changes in user density, FoV, or room layout). However, it is evident from the results in Table 3 that the proposed deep learning model offers a much-optimized performance compared with the traditional method of [23]. Specifically, the KDR is improved from 41% in [23] to only 14% in the proposed deep learning method, hence, improving the KGR from 5 bps to 79 bps. This performance is attributed to the deep learning-based optimized quantization levels. It should be highlighted that both methods fail passing the NIST randomness tests when subject to dynamic setups (i.e., with changes in user density, FoV, or room layout). Hence, we propose next the ensemble method. Nevertheless, this comparison motivates the need for a deep learning-based approach for secret key generation.

IV. ROBUST DEEP LEARNING BASED KEY GENERATION STRATEGY

This section first analyses the reasons behind the deteriorating performance and limited generalization ability of the deep LSTM models in dynamic environments as shown in Table 2. Then, we present an online learning strategy that overcomes such a limitation.

A. CONCEPT DRIFT ANALYSIS

To provide a comprehensive analysis, we study three key metrics: (a) *Evolution of the average CIR with time*, this metric allows us to understand how the average CIR fluctuates over the wandering stage, (b) *CDF of average CIR* that is used to give a statistical overview of the CIR values, and (c) *CDF of average outage events* that measures how frequently communication outages occur and helps to understand the CIR statistics. Here, the "average" is computed over the mobility traces under a specific scenario. The results are shown in Figure 6 and we summarize our findings as follows:

- The average CIR varies based on the user density, FoVs, and room layout. The CIR value is highest for cases with a single UE and it decreases as more UEs are wandering in the room for any FoV or room layout. This is due to the higher chance of blockage with the number of users in the room, which leads to link outages and hence deterioration in CIR. Also, above 60° FoV, the CIR remains consistent with a minor increment as the AP covers most of the room. The average UL and DL CIRs differ, this is due to the distinct IR and VL characteristics resulting in channel non-ideal reciprocity. The CIR characteristics are also specific to the room layout as it is attributed to the interactions among the users and the environment.
- In scenarios with a low number of users ($1-6$ UE), the CIR CDFs are distinct regardless of the FoV or the layout of the room. However, when the user density is high and close ($6-8$ UE), the CDFs tend to cluster. This suggests that as the number of users increases, the impact on the CIR becomes more uniform across different settings.
- The variations in the CIR CDFs with user density can be understood from the link outage behavior. The distribution of the outage rates varies with the number of users. Higher user density tends to cause high blockage events, and thus higher outage rates, which impacts CIR. It should be noted that increasing the FoV from 30°

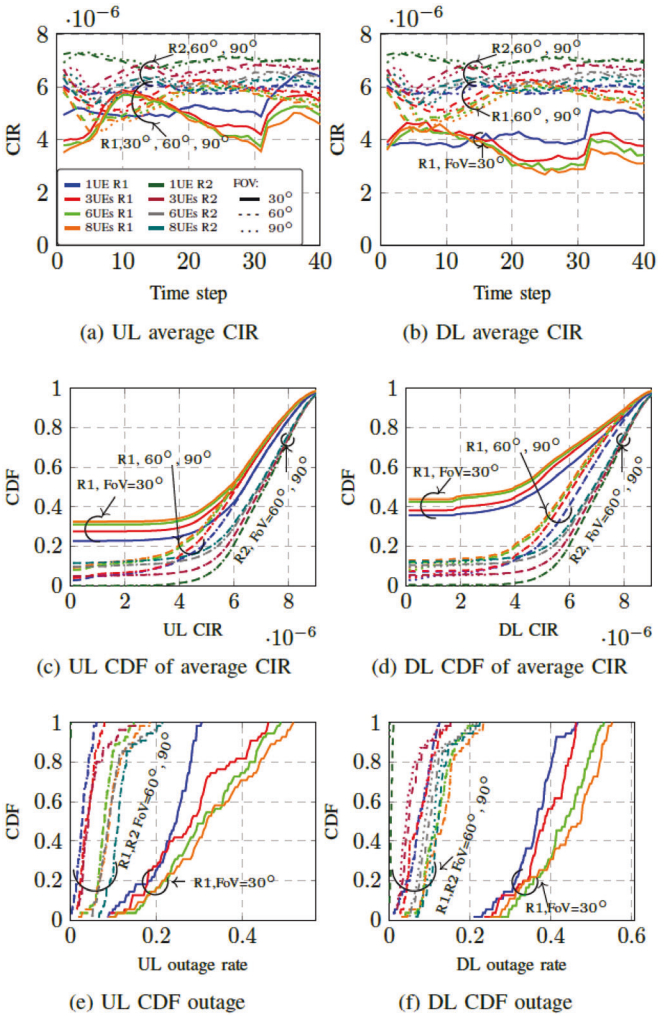


FIGURE 6. The average results are evaluated over 1, 000 distinct mobility traces for scenarios with 1, 3, 6, and 8 users in R1 and R2 layouts. The FoVs considered are 30°, 60°, and 90°. Sub-figures (a) and (b) depict the temporal evolution of CIR for the UL and DL channels during the wandering stage, respectively. CDFs for the UL and DL average CIRs are shown in sub-figures (c) and (d), while sub-figures (e) and (f) illustrate the CDFs for the average outage rates for UL and DL.

to 60° significantly reduces the outage rate, suggesting that a wider FoV experiences channel characteristics.

The shifts that occur in the CIR CDFs due to changes in the user densities, room layouts, and FoVs are known as concept drifts. These drifts are what causing the deterioration in the LSTM model performance summarized in Table 2. Formally, concept drift can be described using the Bayes posterior probability of a class given a specific instance, denoted as $P(y|x)$, where $y \in \mathbf{Y}$ represents a class label and $x \in \mathbf{X}$ denotes a normalized CIR instance. This probability is determined by the likelihood $P(x|y)$, prior probability $P(y)$, and evidence $P(x)$. concept drift occurs when the posterior probability changes over time, i.e., $P_{t+1}(y|x) \neq P_t(y|x)$ [50], [51], [52]. In practical scenarios within dynamic indoor systems where the channel distribution changes due to fluctuations in the number of users, FoV setting, and room layouts, concept drift becomes more probable, as shown in Figure 6. This concept drift presents a challenge from a data-driven perspective as traditional statistical assumptions and models assume stationary data distribution. However, when concept drift occurs, these assumptions are violated, leading to performance degradation and inaccurate predictions as summarized in Table 2. Hence, more efforts are needed to address this phenomenon and attain a stable performance for the LSTM model.

B. DATASET SIMILARITY ANALYSIS

This subsection explores the similarity in the probability distributions of CIR under different settings. This analysis will inform our proposed strategy to mitigate concept drift. Dataset similarity refers to the comparison of probability distributions to determine how closely they match or differ. Formally, it assesses the statistical similarity between distributions, helping to quantify the extent of overlap or divergence between datasets. The Kolmogorov-Smirnov (K-S) test is commonly used to measure this similarity, providing a quantitative assessment beyond visual inspection [53]. By comparing test p-values to a threshold α, we can determine if the two datasets are significantly similar or different. The null hypothesis for the K-S test states that two samples are drawn from the same distribution. Rejection of the null hypothesis suggests a significant difference between the samples. A summary of the K-S test is provided in Algorithm 3, where $|\mathbf{H}_1|$ and $|\mathbf{H}_2|$ represent the sizes or number of data points in the two datasets being compared and $\alpha = 0.05$ [53].

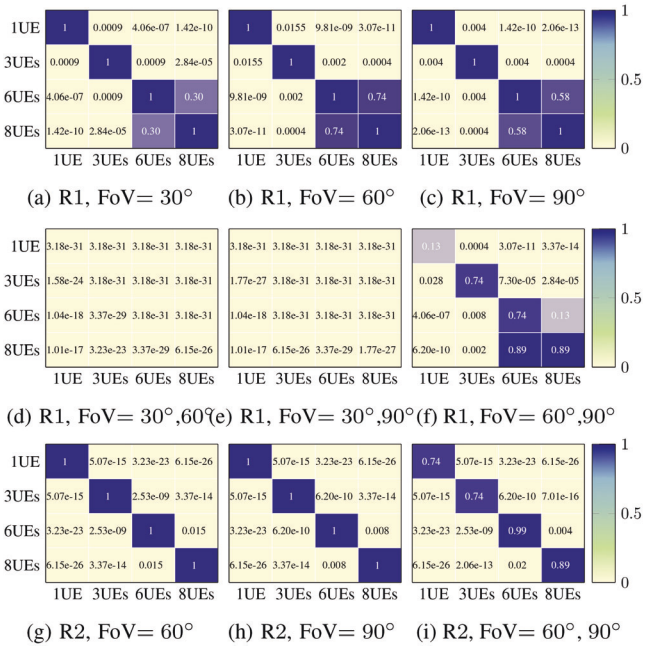
The heat maps in Figure 7 show the results of the K-S tests across 144 comparison of PDF similarity. The heatmap displays p-values calculated using the K-S test. In this context, the test is applied to compare the CDFs of different scenarios with 1, 3, 6, and 8 UEs with different FoVs 30°, 60°, and 90° in room layouts R1 and R2. Specifically, Sub-Figures 5(a)-(c) examine the similarity at FoVs 30°, 60°, and 90°, respectively, in R1 while Figures 5(d)-(f) test the similarity across different FoVs in R1, for instance in 5(f), 0.13 represent the p-value of similarity between 1 UE of 60° and 1 UE of 90°. Likewise, Figures 5(g) and 5(h) represent the different number of users with 60° and 90° FoV in R2, where sub-figure (i) shows the similarity between 60° and 90° FoVs in R2. Areas with high p-values (blue shaded) suggest that the CDFs of those scenarios are more similar, while areas with lower p-values (yellow) indicate more dissimilarity. Hence, the diagonal of the heatmap displays the highest p-value of 1 under the same FoV because it represents the comparison of identical CDFs. This also holds between 60° and 90° FoVs for any room layout. Notably, we observe significant similarities between the scenarios

Algorithm 3 Kolmogorov-Smirnov Test

procedure K-S-TEST($\mathbf{H}_1$, $\mathbf{H}_2$, α)
  Set significance level: α
  Compute empirical CDF for $\mathbf{H}_1$: $F_{\mathbf{H}_1}(x)$
  Compute empirical CDF for $\mathbf{H}_2$: $F_{\mathbf{H}_2}(x)$
  Calculate KS-statistic: $\gamma = \max |F_{\mathbf{H}_1}(x) - F_{\mathbf{H}_2}(x)|$
  Calculate p-value:

$$p = 2\left(1 - \sum_{l=1}^{\infty}(-1)^{l-1}e^{-2l^2\gamma^2}\right) \times \sqrt{\frac{|\mathbf{H}_1| + |\mathbf{H}_2|}{|\mathbf{H}_1| \times |\mathbf{H}_2|}}$$

  if $p < \alpha$ **then**
    Reject the null hypothesis
  else
    Do not reject the null hypothesis
  end if
end procedure



(a) R1, FoV= 30° (b) R1, FoV= 60° (c) R1, FoV= 90°

(d) R1, FoV= 30°,60° (e) R1, FoV= 30°,90° (f) R1, FoV= 60°,90°

(g) R2, FoV= 60° (h) R2, FoV= 90° (i) R2, FoV= 60°, 90°

FIGURE 7. Heatmap representing p-values from K-S test of CIR similarities for different users, FoVs, and room layouts.

involving 6 UEs and 8 UEs for any FoV in R1, which is not the case in R2. We can conclude that scenarios involving 6 UEs and 8 UEs are statistically indistinguishable in R1. Although in R2, there is some resemblance between the CDFs of 6 UEs and 8 UEs, they do not pass the K-S test and are, therefore, not considered similar. Hence, the similarity among users within a specific FoV is affected by the room layout. On the other hand, there is a significant resemblance between the 60° and 90° FoV in both R1 and R2. Consequently, we can infer that for FoV values above 60°, all the statistics exhibit similarity.

In this study, distribution similarity has been used to assess whether the channel evolution across different scenarios is statistically similar. If a distributional similarity is established between scenarios, we can combine the channel data resulting from such scenarios and develop a joint deep learning model encompassing such scenarios, hence, enhancing the generalizability and robustness of the model without needing to address each scenario separately, as will be presented in the next sub-section.

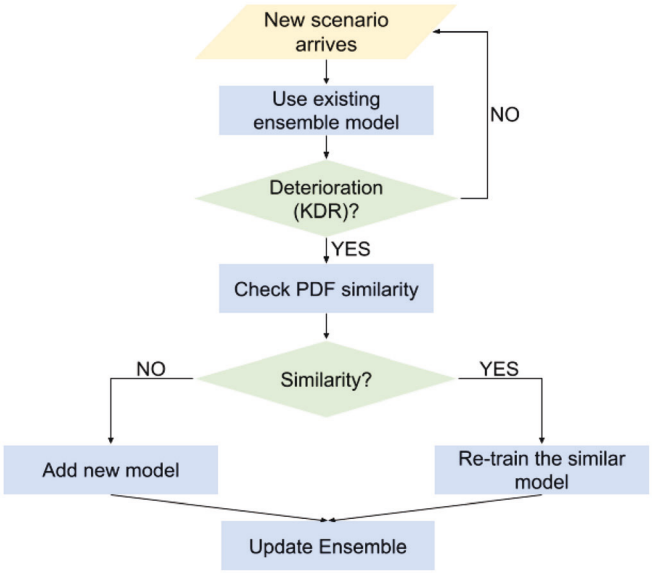


FIGURE 8. Illustration of the proposed online ensemble strategy.

C. CONCEPT DRIFT AWARE ONLINE LEARNING STRATEGY

To address the impact of concept drift and maintain stable performance, an ensemble approach is proposed. Table 2 shows that models trained and tested on identical user densities exhibit superior performance. Thus, combining these custom models in an ensemble manner is expected to enhance the generalization ability. However, using too many custom models in the ensemble strategy is too complex. Instead, we propose to benefit from the similarity of some distributions based on the K-S test to combine some models and reduce the complexity of the ensemble strategy. The following remarks can be made based on the similarity analysis of Section IV-B:

- In R1, it is possible to combine the custom models representing cases with 6 and 8 UEs.
- Beyond a FoV of 60°, there is no need for separate models as the distribution remains consistent.
- R1 and R2 layouts have distinct distributions and as such would need their separate models.

Accordingly, we can reduce the number of required custom models in R1 from 9 to 6, which represents a 33% reduction in complexity. These would make the base custom models in the ensemble strategy. For any incoming CIR data, we will first evaluate them against the existing (base) models. The model performance, represented by KDR, is monitored to identify any deterioration (drift detection). If such deterioration is observed, the online strategy checks the similarity of the incoming data distribution with those of the base models. Depending on the similarity, the strategy takes one of two actions: it either updates the existing model that is most similar in distribution or trains a new model if no similarity is found. This online process ensures that the ensemble model remains robust and updated according to any detected drifts, thereby enhancing its generalization capabilities across various scenarios, while reducing the complexity. The illustration of the proposed online strategy is presented in Figure 8.

For the ensemble models, we examine the following techniques.

1) BAGGING ENSEMBLE

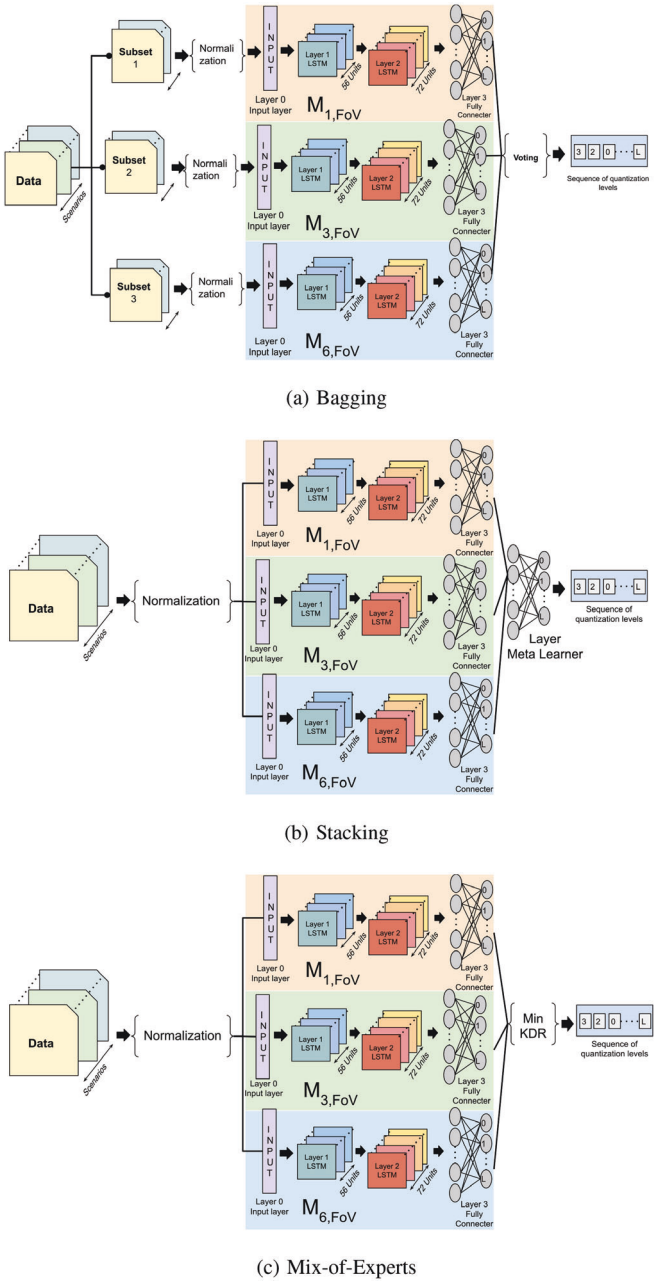
This technique is designed to enhance the consistency and accuracy of ML models used in statistical classification. In this technique, we initially create multiple subsets from the original dataset ($\mathbf{X}$, $\mathbf{Y}$), then proceed to train models $\mathbf{M}_1$, $\mathbf{M}_3$, and $\mathbf{M}_6$ on each subset, and finally aggregate their decisions. The ultimate output is determined by a majority vote for the minimum KDR. This technique serves to reduce model variance and alleviate the risk of overfitting. The bagging ensemble strategy investigated in our study is illustrated in Figure 9(a).

2) STACKING ENSEMBLE

In this technique, models $\mathbf{M}_1$, $\mathbf{M}_3$, and $\mathbf{M}_6$ are trained separately, as shown in Figure 9(b). When dealing with testing data, the results produced by these main models are used as inputs for another layer, called a meta-learner. This meta-learner is taught to combine the decisions of the individual models to make a final decision. The meta-learner follows the same structure explained in Section III-D.

3) MIX-OF-EXPERTS ENSEMBLE

In this technique, we employ a mix-of-experts (MoEs) strategy to reach the best decision based on individual models. The models $\mathbf{M}_1$, $\mathbf{M}_3$, and $\mathbf{M}_6$ are each trained on statistically distinct datasets and fine-tuned independently. When new testing data is presented, the final decision is made by selecting the class with the lowest KDR. Each 'expert' model thus has a special understanding of a subset



(a) Bagging

(b) Stacking

(c) Mix-of-Experts

FIGURE 9. Illustration of the investigated ensemble techniques.

of the data, and the model that provides the lowest KDR is selected as the authoritative expert for that specific test instance. Figure 9(c) illustrates the MoE technique. To highlight the difference between bagging and MoE. In bagging, the data is initially partitioned into subsets. Individual models are tested on these subsets and then used to vote on the output for each instance. The final prediction is made by aggregating these votes, if there is a tie we use a random selection method. The sequence of predictions is collated to form the complete output. In contrast, MoEs employ the entire dataset to test, with each model generating a complete sequence of predictions. The decision on which model's sequence to choose is guided by the KDR, the

TABLE 4. Summary of ensemble KDR, KGR and NIST. The results for 6 and 8 UEs are grouped in a single column due to the high similarity in their CIR characteristics.

	KDR%			KGR(bits/s)			NIST		
Ensemble	1UE	3UEs	6/8UEs	1UE	3UEs	6/8UEs	1UE	3UEs	6/8UEs
Bag$_{30^\circ}$	42.3	41.3	37.9	50.5	11.7	16.3	Failed	Failed	Failed
Stack$_{30^\circ}$	58.9	54.1	50.2	30.8	35.2	37.8	Failed	Failed	Failed
MoEs$_{30^\circ}$	**14.3**	**15.2**	**15.7**	**79.1**	**77.8**	**77.4**	**Passed**	**Passed**	**Passed**
Bag$_{60^\circ,90^\circ}$	37.1	34.6	35.8	32.1	10.1	14.2	Failed	Failed	Failed
Stack$_{60^\circ,90^\circ}$	70.1	70.6	70.1	15.4	15.2	15.8	Failed	Failed	Failed
MoEs$_{60^\circ,90^\circ}$	**7.9**	**8.4**	**8.5**	**89.8**	**88.9**	**88.9**	**Passed**	**Passed**	**Passed**

model with the lowest KDR for a given sequence is the most accurate for that instance, and its sequence is selected.

Table 4 compares the KDR, KGR, and NIST of the three ensemble strategies. The results are summarized next:

- The MoEs ensemble model outperforms the bagging and stacking ensemble techniques and offers the minimum KDR and maximum KGR with a stable performance, showing only 1% performance deterioration under any FoV and user density.
- The MoEs ensemble model is the only one that generates keys passing all the NIST randomness tests. This means the keys are random and safe to use, no matter how many users are around and whatever FoV is used.
- In comparing Table 2 and Table 4, the MoEs model presented in Table 4 offers robust performance against dynamic setups involving various user density levels, achieving consistently lower KDR and higher KGR values compared to the custom and general density models presented in Table 2. For instance, model $\mathbf{M}_{\mathrm{G},30^\circ}$ in Table 2 offers KDR and KGR that vary between $38-44\%$ and $48-43$ bps, respectively, depending on the user density level, and fails all NIST tests. Similarly, other custom models in Table 2 offer KDR and KGR that vary between $14-43\%$ and $40-79$ bps, respectively, and sometimes pass or fail the NIST tests, depending on the user density level. On the other hand, the MoEs model in Table 4 maintains a more stable KDR and KGR of $14-15\%$ and $77-79$ bps, respectively, while always passing the NIST tests, regardless of the user density level.
- When it comes to the proposed robust online strategy, the models constituting the MoEs ensemble run in parallel, hence, maintaining the time complexity the same as the deep strategy of Section III. As for the space complexity, while the space complexity grows linearly with the number of models constituting the

MoE ensemble, the proposed clustering method based on PDF similarity reduces the number of models needed by 50% in the investigated case studies. Also, additional models are needed only if a deviation in the KDR is observed.

V. CONCLUSION AND FUTURE WORK

In this paper, we studied WSKG in multi-user dynamic LiFi networks. We introduced a novel approach utilizing an LSTM model within a deep learning framework to optimize the downlink quantization thresholds at the AP, aiming to generate preliminary keys with minimal KDR in both uplink and downlink. Our analysis showed that dynamic scenarios with different numbers of users, FoVs, and room layouts introduce concept drifts, resulting in limited generalization ability for the LSTM model and deteriorating the performance by $28-44\%$. Further analysis of distribution similarity based on K-S tests suggested that some scenarios exhibit similar distributions and thus their corresponding models can be combined. This finding allowed us to develop an online ensemble strategy that relies on fewer custom (base) models to maintain a stable performance. The custom models can be updated whenever a new scenario is found to have a distinct distribution from the base models. Our results showed that an ensemble strategy based on the MoE technique results in a stable performance with minimum KDR and maximum KGR deviating only by 1% under any user density or FoV.

Given the impact of environmental factors and the dynamicity of user density, transfer learning could be beneficial within a single-room setup, though it may present challenges across different room setups. Nonetheless, transfer learning holds strong potential for addressing the challenges outlined in this work. Hence, our future work will consider the adoption of transfer learning to further enhance the robustness of WSKG in dynamic environments.

In this work, we assumed there were no attacks. In future work, we will assess the risk of passive attacks, exploring how an eavesdropper can compromise the key between the legitimate user and the AP. Furthermore, we will investigate the implications of active attacks, where an attacker actively interferes with the communication process to deceive the system. Our future research will enhance the security framework by incorporating advanced attack detection mechanisms and developing robust countermeasures to protect against passive and active threats. By addressing these challenges, we aim to further strengthen the resilience and reliability of WSKG in dynamic LiFi networks.

ACKNOWLEDGMENT

Dataset generation was supported by the TNTech HPC cluster funded by NSF award 2127188.

REFERENCES

[1] A. N. Kadhim and S. B. Sadkhan, "Security threats in wireless network communication-status, challenges, and future trends," in *Proc. Int. Conf. Adv. Comput. Appl. (ACA)*, 2021, pp. 176–181.

[2] C. Paar and J. Pelzl, *Understanding Cryptography: A Textbook for Students and Practitioners*, 1st ed. Berlin, Germany: Springer, 2009.

[3] T. Q. Duong, "Keynote talk #1: Trusted communications with physical layer security for 5G and beyond," in *Proc. Int. Conf. Adv. Technol. Commun. (ATC)*, 2017, p. 34.

[4] M. Adil, S. Wyne, and S. J. Nawaz, "On quantization for secret key generation from wireless channel samples," *IEEE Access*, vol. 9, pp. 21653–21668, 2021.

[5] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-rate uncorrelated bit extraction for shared secret key generation from channel measurements," *IEEE Trans. Mobile Comput.*, vol. 9, no. 1, pp. 17–30, Jan. 2010.

[6] K. Ren, H. Su, and Q. Wang, "Secret key generation exploiting channel characteristics in wireless communications," *IEEE Wireless Commun.*, vol. 18, no. 4, pp. 6–12, Aug. 2011.

[7] K. Zeng, "Physical layer key generation in wireless networks: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 33–39, Jun. 2015.

[8] M. Bottarelli, P. Karadimas, G. Epiphaniou, D. K. B. Ismail, and C. Maple, "Adaptive and optimum secret key establishment for secure vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2310–2321, Mar. 2021.

[9] S. Ribouh, K. Phan, A. V. Malawade, Y. Elhillali, A. Rivenq, and M. A. A. Faruque, "Channel state information-based cryptographic key generation for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7496–7507, Dec. 2021.

[10] E. O. Torshizi and W. Henkel, "Exploiting FDD channel reciprocity for physical layer secret key generation in IoT networks," *IEEE Wireless Commun. Lett.*, vol. 28, no. 6, pp. 1268–1272, Jun. 2024.

[11] (Cisco, San Jose, CA, USA). *Cisco Vision: 5G—Thriving Indoors*. [Online]. Available: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/ultra-services-platform/5g-ran-indoor.pdf

[12] X. Wu, M. D. Soltani, L. Zhou, M. Safari, and H. Haas, "Hybrid LiFi and WiFi networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1398–1420, 2nd Quart., 2021.

[13] Z.-Y. Wu, M. Ismail, J. Kong, E. Serpedin, and J. Wang, "Channel characterization and realization of mobile optical wireless communications," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6426–6439, Oct. 2020.

[14] W. Xu, J. Zhang, S. Huang, C. Luo, and W. Li, "Key generation for Internet of Things: A contemporary survey," *ACM Comput. Surv.*, vol. 54, no. 1, p. 14, 2021.

[15] A. I. Sulimov, A. A. Galiev, A. V. Karpov, and V. V. Markelov, "Verification of wireless key generation using software defined radio," in *Proc. Int. Siberian Conf. Control Commun. (SIBCON)*, 2019, pp. 1–6.

[16] J. Huang and T. Jiang, "Dynamic secret key generation exploiting ultra-wideband wireless channel characteristics," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2015, pp. 1701–1706.

[17] L. Jiao, J. Tang, and K. Zeng, "Physical layer key generation using virtual AoA and AoD of mmWave massive MIMO channel," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, 2018, pp. 1–9.

[18] N. Felkaroski and M. Petri, "Secret key generation based on channel state information in a mmWave communication system," in *Proc. 12th Int. ITG Conf. Syst., Commun. Coding (SCC)*, 2019, pp. 1–6.

[19] H. Hentilä, Y. Y. Shkel, and V. Koivunen, "Secret key generation using short blocklength polar coding over wireless channels," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 1, pp. 144–157, Jan. 2022.

[20] Z. Ji et al., "Wireless secret key generation for distributed antenna systems: A joint space-time-frequency perspective," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 633–647, Jan. 2022.

[21] L. Jiao et al., "Efficient physical layer group key generation in 5G wireless networks," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, 2020, pp. 1–9.

[22] Y. Yang, M. Ma, S. Aïssa, and L. Hanzo, "Physical-layer secret key generation via CQI-mapped spatial modulation in multi-hop wiretap ad-hoc networks," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1322–1334, 2021.

[23] E. Mahalal, M. Ismail, Z. Wu, and M. M. Fouda, "Characterization of secret key generation in 5G+ indoor mobile LiFi networks," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, 2022, pp. 1–6.

[24] Y. M. Al-Moliki, M. T. Alresheedi, and Y. Al-Harthi, "Secret key generation protocol for optical OFDM systems in indoor VLC networks," *IEEE Photon. J.*, vol. 9, no. 2, pp. 1–15, Apr. 2017.

[25] Y. Gan, X. Lei, Y. Xiao, H. Hou, and X. Zhou, "Improved channel information extraction toward efficient secret key generation," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, 2019, pp. 97–101.
[26] J. W. Wallace and R. K. Sharma, "Automatic secret keys from reciprocal MIMO wireless channels: Measurement and analysis," *IEEE Trans. Inf. Forensics Security*, vol. 5, pp. 381–392, 2010.
[27] H. Jin, K. Huang, L. Jin, Z. Zhong, and Y. Chen, "Physical-layer secret key generation with correlated eavesdropping channel," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, 2018, pp. 226–231.
[28] L. Wang, H. An, H. Zhu, and W. Liu, "MobiKey: Mobility-based secret key generation in smart home," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7590–7600, Aug. 2020.
[29] L. Peng, G. Li, and A. Hu, "Channel reciprocity improvement of secret key generation with loop-back transmissions," in *Proc. IEEE 17th Int. Conf. Commun. Technol. (ICCT)*, 2017, pp. 193–198.
[30] M. Yuliana, Wirawan, Suwadi, Endroyono, and T. Suryani, "Enhancing channel reciprocity of secret key generation scheme by using modified polynomial regression method," in *Proc. Int. Conf. Comput. Eng., Netw. Intell. Multimedia (CENIM)*, 2018, pp. 35–40.
[31] C. Feng and L. Sun, "Physical layer key generation from wireless channels with non-ideal channel reciprocity: A deep learning based approach," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC)*, 2022, pp. 1–6.
[32] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "Deep learning for hardware-impaired wireless secret key generation with man-in-the-middle attacks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.
[33] L. Jiao, G. Sun, J. Le, and K. Zeng, "Machine learning-assisted wireless PHY key generation with reconfigurable intelligent surfaces," in *Proc. 3rd ACM Workshop Wireless Secur. Mach. Learn.*, 2021, pp. 61–66.
[34] X. Wei and D. Saha, "KNEW: Key generation using neural networks from wireless channels," in *Proc. ACM Workshop Wireless Secur. Mach. Learn.*, 2022, pp. 45–50.
[35] X. Zhang, G. Li, J. Zhang, A. Hu, Z. Hou, and B. Xiao, "Deep-learning-based physical-layer secret key generation for FDD systems," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 6081–6094, Apr. 2022.
[36] Z. Hou and X. Zhang, "Secret key generation scheme based on generative adversarial networks in FDD systems," in *Proc. INFOCOM IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.
[37] R. Zhao, Q. Qin, N. Xu, G. Nan, Q. Cui, and X. Tao, "SemKey: Boosting secret key generation for RIS-assisted semantic communication systems," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC)*, 2022, pp. 1–5.
[38] J. Han, Y. Zhou, G. Liu, T. Liu, and X. Zeng, "A novel physical layer key generation method based on WGAN-GP adversarial autoencoder," in *Proc. 4th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, 2022, pp. 1–6.
[39] E. Mahalal, M. Ismail, Z.-Y. Wu, M. M. Fouda, Z. M. Fadlullah, and N. Kato, "Robust deep learning-based secret key generation in dynamic LiFi networks against concept drift," in *Proc. IEEE 21st Consum. Commun. Netw. Conf. (CCNC)*, 2024, pp. 899–904.
[40] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008.
[41] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, Jun. 2011.
[42] M. D. Soltani, A. A. Purwita, Z. Zeng, H. Haas, and M. Safari, "Modeling the random orientation of mobile devices: Measurement, analysis and LiFi use case," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2157–2172, Mar. 2019.
[43] P. Chvojka, S. Zvanovec, P. A. Haigh, and Z. Ghassemlooy, "Channel characteristics of visible light communications within dynamic indoor environment," *J. Lightw. Technol.*, vol. 33, no. 9, pp. 1719–1725, May 1, 2015.
[44] K. Lee, H. Park, and J. R. Barry, "Indoor channel characteristics for visible light communications," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 217–219, Feb. 2011.
[45] J. Zhang, G. Li, A. Marshall, A. Hu, and L. Hanzo, "A new frontier for IoT security emerging from three decades of key generation relying on wireless channels," *IEEE Access*, vol. 8, pp. 138406–138446, 2020.
[46] (Nat. Inst. Stand. Technol., Gaithersburg, MD, USA). *Secure Hash Standard (SHS). U.S. Department of Commerce*. 2015. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf
[47] K. Liu, S. Primak, and X. Wang, "On secret key generation from multiple observations of wireless channels," in *Proc. IEEE Int. Conf. Commun. Syst.*, 2014, pp. 147–151.
[48] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, 2008, pp. 128–139.
[49] L. Bassham et al., "A statistical test suite for random and pseudo-random number generators for cryptographic applications," Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, document NIST SP 800-22, 2010, Accessed: Jun. 14, 2023. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906762
[50] R. Elwell and R. Polikar, "Incremental learning of concept drift in nonstationary environments," *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1517–1531, Oct. 2011.
[51] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2346–2363, Dec. 2019.
[52] D. Brzezinski and J. Stefanowski, "Reacting to different types of concept drift: The accuracy updated ensemble algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 81–94, Jan. 2014.
[53] J. W. Conover, *Practical Nonparametric Statistics 3rd Ed*. Hoboken, NJ, USA: Wiley, 2006. [Online]. Available: https://books.google.com/books?id=UBV2VwCxrMcC

ELMAHEDI MAHALAL (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications from the University of Science and Technology Houari Boumediene, Algeria, in 2017 and 2019, respectively, and the Ph.D. degree in computer science from Tennessee Technological University, USA, in 2024. He is currently an Assistant Professor of Cybersecurity and Networks with the Department of Computer Science, University of New Haven, West Haven, CT, USA. His research interests include wireless networks, artificial intelligence, and security.

ESLAM HASAN (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering (electronics and communications) from Mansoura University, Cairo, Egypt, in 2013 and 2020, respectively. He is currently pursuing the Ph.D. degree with the Computer Science Department, Faculty of Engineering, Tennessee Tech University, Cookeville, TN, USA. He is an experienced academic with ten years of research and teaching experience in prestigious institutes, including Tennessee Tech University, The American University in Cairo, and Mansoura University. His research focuses on AI, cybersecurity, and 5G+ wireless networks. He has served as a reviewer for several IEEE conferences. He served as the Session Chair for three sessions in the IEEE VTC Fall 2024. He served as a TPC Member for IEEE VCC 2024.

MUHAMMAD ISMAIL (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering (electronics and communications) from Ain Shams University, Cairo, Egypt, in 2007 and 2009, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2013. He is the Director of the Cybersecurity Education, Research, and Outreach Center, and an Associate Professor with the Department of Computer Science, Tennessee Technological University, Cookeville, TN, USA. He was a co-recipient of the Best Paper Awards in the IEEE ICC 2014, the IEEE GLOBECOM 2014, the SGRE 2015 and 2024, the Green 2016, and the IEEE IS 2020, and the Best Conference Paper Award from the IEEE Communications Society Technical Committee on Green Communications and Networking for his publication in IEEE ICC 2019. He was the Track Chair in the IEEE Globecom 2024, the Track Co-Chair in the IEEE SmartGridComm 2023, the Workshop Co-Chair of the IEEE Greencom 2018, the Track Co-Chair of the IEEE VTC 2017 and 2016, the Publicity and Publication Co-Chair of the CROWNCOM 2015, and the Web-Chair of the IEEE INFOCOM 2014. He was an Associate Editor of the *IET Communications*, *Physical Communication*, and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He was an Editorial Assistant of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2011 to 2013. He is an Associate Editor of the IEEE INTERNET OF THINGS JOURNAL and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He has been a technical reviewer of several IEEE conferences and journals.

ZI-YANG WU (Member, IEEE) received the B.S. degree in electronic science and technology, the M.S. degree in circuits and systems, and the Ph.D. degree in control science and engineering from Northeastern University, Shenyang, China, in 2014, 2016, and 2020, respectively. He was also a joint Ph.D. student with the Department of Electrical and Computer Engineering, Texas A&M University at College Station, College Station, USA, from 2018 to 2019. He is currently an Associate Researcher with the College of Information Science and Engineering, Northeastern University. His research focuses on wireless communications and ML. He was a recipient of the Best Paper Awards in the IEEE International Conference on Intelligent Systems in 2020. He has served as a reviewer for several IEEE journals.

MOSTAFA M. FOUDA (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Benha University, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in information sciences from Tohoku University, Japan, in 2011. He is an Associate Professor with the Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA. He is also a Full Professor with Benha University. He was an Assistant Professor with Tohoku University and a Postdoctoral Research Associate with Tennessee Technological University, Cookeville, TN, USA. He has (co-)authored more than 280 technical publications. He has received several research grants, including NSF Japan-U.S. Network Opportunity 3. He has guest-edited several special issues covering various emerging topics in communications, networking, and health analytics. His current research focuses on cybersecurity, communication networks, signal processing, wireless mobile communications, smart healthcare, smart grids, AI, and IoT. He serves on the editorial board of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, and IEEE ACCESS.

ZUBAIR MD FADLULLAH (Senior Member, IEEE) is currently an Associate Professor with the Computer Science Department, University of Western Ontario, Canada. He was previously the Smart Health Technology Research Chair with the Thunder Bay Regional Health Research Institute, and an Associate Professor with Lakehead University, Canada. He was an Associate Professor with the Graduate School of Information Sciences (GSIS), Tohoku University, Japan, from 2017 to 2019. He also served with GSIS as an Assistant Professor from 2011 to 2017. His main research interests are in the areas of emerging communication networks and interdisciplinary domains such as 5G and beyond (B5G) networks, Internet of Things-based smart health technology and medical analytics, and application of AI and optimization methods for solving computer science and communication system problems. He was a recipient of the prestigious Dean's and President's Awards from Tohoku University in March 2011, and the IEEE Asia Pacific Outstanding Researcher Award in 2015 and the NEC Tokin Award for research in 2016, for his outstanding contributions. He has also received several best paper awards at conferences, including IWCMC, Globecom, and ICNIDC. He received the Lakehead University Research Excellence Award in February 2022. He is also an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

NEI KATO (Fellow, IEEE) is a Full Professor and the Dean of the Graduate School of Information Sciences, Tohoku University. He has researched on computer networking, wireless mobile communications, satellite communications, ad hoc and sensor and mesh networks, AAV networks, AI, IoT, and big data. He is the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, and the Fellow Committee Chair of IEEE VTS. He is a Fellow of the Engineering Academy of Japan and of IEICE.