DroneAudioID: A Lightweight Acoustic Fingerprint-Based Drone Authentication System for Secure Drone Delivery

Meng Zhang, Li Lu, *Member, IEEE*, Yuhan Wu, Zheng Yan, Jiaqi Sun, Feng Lin, *Senior Member, IEEE*, and Kui Ren, *Fellow, IEEE*

***Abstract*—With the increasing accessibility of drones, they have been warmly embraced across various sectors, especially in low-altitude logistics transportation. However, during drone delivery, legal drones dispatched by logistics companies are susceptible to malicious attacks, resulting in package theft or substitution. To address this, existing works focus on designing drone authentication to secure drone delivery. However, most of these methods require expensive specialized equipment, such as high-quality microphones and professional recording devices, resulting in high real-world application costs. In this paper, we propose *DroneAudioID*, a lightweight acoustic fingerprint-based drone authentication system that relies solely on common mobile devices. The basic idea is to employ acoustic fingerprints to authenticate different drones of the same model based on differences in fundamental frequency and harmonic components of drone audio. Specifically, the drone audio is recorded by a mobile device instead of sophisticated equipment. We apply wavelet transform to remove high-frequency noise during data preprocessing. Then, specialized filter banks are designed for feature extraction, leveraging the frequency characteristics of drone audio. Finally, we construct a Bi-Long Short-Term Memory (Bi-LSTM) with an Open-Max model for open-set classification. Extensive experiments are conducted on eight crafts drones of *DJIMini*2, showing an authentication accuracy of 99.6%. A series of comprehensive experiments further validate *DroneAudioID*'s capability to defend against various attacks.**

***Index Terms*—Drone authentication, acoustic fingerprint, Bi-LSTM, secure drone delivery.**

I. Introduction

RECENT years have witnessed the drone industry ushering in rapid development. It has become one of the important development directions with a market projection of $29 billion within the next decade [1]. Amidst the COVID-19 outbreak, a number of countries have utilized drones as a means to offer secure and contact-free package transportation in hard-to-reach areas. This approach has proven to be more robust and reliable during these challenging times. Low-altitude logistics delivery refers to the use of drones to deliver packages to designated locations, with the advantages of autonomy, effectiveness, and timeliness. Leading players in online retailers (e.g., Amazon and Alibaba) are prototyping systems to deliver packages from their warehouses to customers via drones [2]. Delivery companies, such as DHL and EMS, are also exploring drones to enhance their speedy delivery options except for traditional transportation methods.

With the soaring popularity of drone delivery, authentication is crucial to prevent drones from accessing resources they are not authorized to access. In the scenario of utilizing drones for delivery services, the drone is responsible for picking up the packages from the sender. There exists an impersonation attack where an unauthorized drone poses as the legal one in order to steal the package. Moreover, even if the drone is legal and transports the packages to the designated receiver, it is susceptible to the man-in-the-middle attack, such attacks could result in the desired packages being substituted during delivery. Furthermore, the need for drone authentication extends beyond package delivery. For example, in surveillance operations, unauthorized drones can pose significant security risks by falsifying their identity to access restricted areas or data. In emergency response scenarios, such as search-and-rescue missions or medical supply drops, confirming the identity of the drone is paramount to ensuring that critical resources are delivered to the right location and are not intercepted by malicious parties. Without authentication, the integrity and reliability of the collected data could be compromised, leading to misleading insights and potentially harmful actions.

For authentication purposes, many drones have software-level digital certificates to indicate the identity of each drone, such as the engraving of the registration number on the drone's exterior and broadcasting unencrypted identity information [3]. As a countermeasure against potential copy attacks, many studies have suggested the adoption of Public Key Infrastructure (PKI) in the context of drones. This approach involves harnessing digital certificates to establish publicly verifiable

Received 10 June 2024; revised 4 November 2024 and 8 December 2024; accepted 3 January 2025. Date of publication 13 January 2025; date of current version 31 January 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFB3107402; in part by the National Natural Science Foundation of China under Grant 62102354, Grant 62032021, and Grant 62372406; and in part by Zhejiang Provincial Natural Science Foundation of China under Grant LY24F020007. The associate editor coordinating the review of this article and approving it for publication was Prof. Kwok-Yan Lam. *(Corresponding author: Li Lu.)*

The authors were with the State Key Laboratory of Blockchain and Data Security, the School of Cyber Science and Technology, and the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China, and also with Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, Hangzhou, Zhejiang 310012, China (e-mail: zhangmengyang@zju.edu.cn; li.lu@zju.edu.cn; yuhan.wu@zju.edu.cn; wargreymon@zju.edu.cn; is_qjuan@zju.edu.cn; flin@zju.edu.cn; kuiren@zju.edu.cn).

Digital Object Identifier 10.1109/TIFS.2025.3527814

identities, effectively functioning as a "virtual license plate" for drones. Unfortunately, these software-based remedies still present hazards. There exist a multitude of attacks that compromise certificates by targeting certificate authorities and web servers [4], [5], in addition to producing fraudulent certificates, ultimately leading to successful impersonation.

Moreover, Remote IDentification (RID) technology plays a role in regulating drone operations. These regulations and protocols mandate drones to periodically transmit telemetry data (latitude, longitude, speed, etc.) to facilitate precise identification and tracking of the drone. However, there is an inherent vulnerability in the RID receiver, it is easily susceptible to high-power spoofing attacks leveraging open drone ID messages. Such attacks can force the receiver into discarding genuine RID transmissions. Exploiting this vulnerability enables attackers to manipulate the drone's RID information, thereby circumventing existing security measures and potentially engaging in nefarious activities.

Except for software-level solutions, the mechanical structure of the drone determining its physical characteristics is inherently unique. Hence, we turn to exploring such characteristics for drone authentication. When drones are in flight, the operation of brush-less motors and propellers generates distinct audio signals. Due to slight manufacturing imperfections, each drone exhibits a unique acoustic fingerprint. By analyzing these acoustic characteristics, we can establish a robust drone authentication system. Acoustic-based methods offer several advantages: they are cost-effective, easy to deploy, and without the requirement of visual or communication signals. This makes them particularly effective compared to other methods, such as radar, vision, or Radio Frequency (RF), which struggle to differentiate between drones of the same model, as these typically share identical size, color, and material.

Existing researches on acoustic-based drone authentication methods have several limitations. First, there is a dependency on specialized equipment. Many existing studies require expensive specialized equipment, such as high-quality microphones and recording devices, making real-world applications challenging. A low-cost, portable authentication scheme is needed. Second, recording drone audio in a studio. Most research is conducted in indoor environments, limiting its application in drone delivery. Drones in outdoor environments may be affected by factors such as wind, traffic noise, etc., leading to changes in their audio characteristics. Thus, an optimal denoising and audio feature extraction method should be proposed. Also, there are security risks. The drone audio is susceptible to being secretly recorded or forged by malicious third parties, leading to security vulnerabilities. A series of comprehensive experiments should be conducted to validate the security performance of the authentication system based on drone audio.

In this paper, we develop *DroneAudioID*, which is a lightweight acoustic fingerprint-based drone authentication system to enhance the security of drone delivery. More specifically, we make the following contributions:

- **A lightweight authentication system is developed.** *DroneAudioID* can enable the recording of drone audio using a mobile device, eliminating the need for elaborate and costly equipment, as well as any necessity for additional deployment of supplementary gear.
- **A novel audio feature extraction scheme is designed specifically for distinguishing the different drones of the same model.** Initially, we utilize wavelet transform to remove unstable high-frequency noises from the recorded drone audio in an outdoor environment. Subsequently, we develop a specific filter bank to extract the fundamental frequency and harmonics components of drone audio as features, which are the key features for distinguishing different drones in the same model.
- **A classifier with an open-set method is structured.** We design a Bi-Long Short-Term Memory(Bi-LSTM) model with the Open-Max method as the classifier. Which can not only authenticate the registered drones but also authenticate the drones that have never been seen before.
- **Comprehensive experiments are conducted to verify the security performance of the authentication system.** *DroneAduioID* can authenticate the drone's legitimacy to defend against impersonation attacks when the sender is shipping the packages. Additionally, the receiver can utilize *DroneAduioID* to determine whether the packages have been substituted during drone delivery to mitigate man-in-the-middle attacks. Moreover, comprehensive experiments have been conducted to verify the ability of *DroneAduioID* to resist impersonation and third-party attacks.

Fig. 1 illustrates the protocol of *DroneAduioID* for drone authentication. During the shipping phase, the sender initiates a request for a drone from the logistics distribution company's Application Programming Platform (APP). Then, the company dispatches a registered drone to the sender's location to pick up the packages. When the drone reaches the pickup location, the sender records the audio of the drone and submits it to the APP for authentication of its legality. At the receiving phase, the receiver records the audio of the drone with packages. This audio is then sent to the APP for authentication of its consistency, which is compared against those recorded during shipping. If the authentication is successful, the drone hands over the packages to the receiver, and the drone delivery process is over.

The rest of this paper is organized as follows. We first present the related work in Section II. The system model and threat model are introduced in Section III. The characteristics of the drone audio are analyzed in Section IV. And Section V presents the design of the *DroneAudioID* system. We present the performance evaluation of *DroneAudioID* in Section VI. Section VII discusses the limitations and possible extensions of this work. We finally conclude in Section VIII.

II. Related Works

In this section, we review recent studies on acoustic sensing techniques and drone detection and identification methods.

A. Acoustic Sensing

Acoustic sensing has been rapidly developed in recent years, with the omnipresence of mobile and wearable devices and

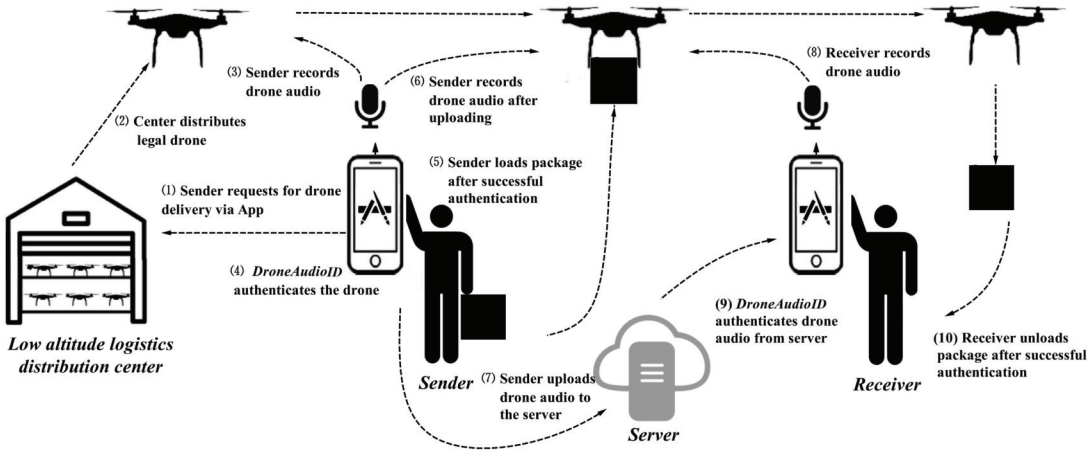


Fig. 1. An exemplary scenario of drone authentication designed for drone delivery.

the advancement in audio chips. Previous studies explored acoustic signals for heartbeat monitoring [6], tracking [7], [8], respiratory symptom detection [9], secret communications [10], [11] and even eye gesture listening [12], etc., which supports various practical applications. With the prevalent applications, recent studies investigate the feasibility of acoustic sensing for security purposes, including user authentication [13], [14], [15], side-channel attacks [16], [17], and even activity eavesdropping [18], [19]. All of these demonstrate the great potential of acoustic sensing in security applications.

B. Drone Detection and Identification

1) Radar-Based Methods: Radar-based methods utilize reflected radar signatures to detect and identify drones. Researches on radar-based detection methods for small commercial drones are explored in references [20], [21]. Schupbach et al. [22] and De Quevedo et al. [23] primarily investigate the detection of drones using specialized equipment, which can detect the drone at a range of 1.2km and 2km, respectively. Although these radar-based methods offer a significantly long detection range, radar equipment is typically large and expensive. More recent works [24], [25], [26] introduce neural networks to realize accurate drone detection on commercial devices, such as mm-wave radars. However, these studies are limited to classifying drone models and do not enable differentiation between individual drones.

2) RF Signature-Based Methods: RF signature-based methods are employed to intercept wireless transmissions between drones and flight controller modules. Passive RF technology [27], [28], [29] plays a crucial role in drone detection and tracking, which involves analyzing signals emitted by drones or fluctuations in signals within the surrounding environment. The following works [30] and [31] investigate the use of raw RF fingerprints for drone-type identification. Additionally, they further confirm the impact of Bluetooth and Wi-Fi signal interference with a series of experiments. RF-based methods require the deployment of an additional set of RF communication systems and are susceptible to environmental interference.

3) Vision-Based Methods: Vision-based methods employ cameras to capture images or videos for drone detection and identification. Mahdavi and Rajabi [32] utilize three classification models for drone identification using fisheye cameras respectively. However, their study solely focuses on drone detection. The classification of different models of drones is addressed by following works [33], [34], [35], which introduce advanced neural networks like Resnet50 and YOLOv4 to identify the category and position. However, the limitation of these studies is the inability to accurately identify the different drones that share the same appearance.

4) Acoustic-Based Methods: Acoustic-based methods leverage the audio generated by drones during flying to make detection and identification, requiring no additional equipment deployment. Early works [36], [37], [38] can only detect the drone existence using acoustic-based methods. The following works [39] and [40] investigate the classification of different models of drones, which validate the effectiveness of using traditional features (e.g., Mel-Frequency Cepstrum Coefficients, MFCC, and Linear Predictive Cepstral Coefficients, LPCC) and deep learning methods (e.g., Long Short-Term Memory, LSTM) to identify drones.

Another work [41] further investigates the possibility of remotely detecting the weight of the payload carried by a commercial drone by analyzing its acoustic fingerprint. However, it requires the use of a professional microphone and a laptop for audio acquisition, which limits its deployment in real-world scenarios. Wu and Zeng [42] develop an audio comparison method that leverages the audio differences between drones and verifier devices for authentication that does not rely on drone audio fingerprints. Other works [43], [44], [45] mainly focus on utilizing other characteristics of drones to facilitate authentication in logistics delivery. Wu et al. [43] propose to sense users' hand waving based on a smartphone's inertial measurement unit while the drone's camera records video. Yang et al. [44] extend this approach, adapting it for authentication in unmanned vehicles.

The acoustic characteristics of drones from the same model are examined in both [46] and [47]. Ramesh et al. [46] delve

into differentiating the audio characteristic of a drone equipped with various motors. They introduce *SoundUAV*, a system that focuses on the differences in the acoustic characteristics of drone motors. They study the audio generated by different motors of a drone, and use MFCC and Support Vector Machine (SVM) to classify the motors, obtaining 99.48% accuracy. However, in the experiment, the drones need to be docked at the specified position, and the motor needs to be dismounted from the drones. The recorded audio solely captures the noise of the operational motor, significantly diminishing its practical application potential. Diao et al. [47] study the classification problem for different drones in the same model. They extract the audio characteristics by the MFCC, delta MFCC and delta-delta MFCC methods respectively, and compared the accuracy with eight different classifiers. The highest accuracy is 96.3%. However, they record the drone audio in an indoor studio with professional microphones, which is limited in drone delivery applications. And the ability of the system to resist attacks was not verified.

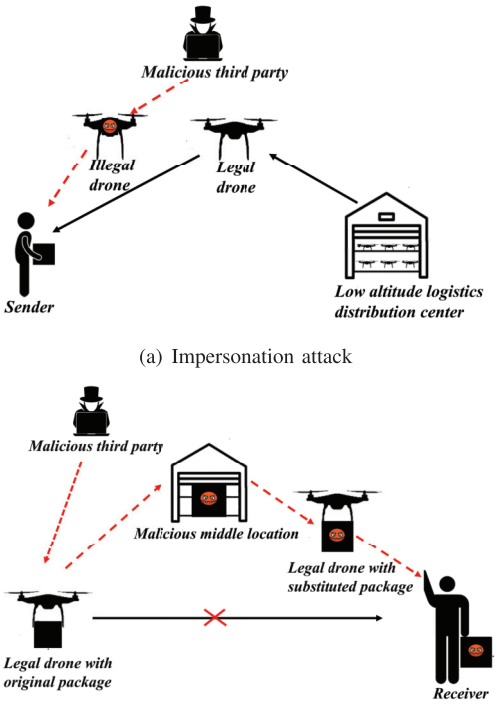
In sum, compared to radar, vision, and RF-based methods, acoustics-based methods offer cost-effectiveness and the ability to distinguish between different drones of the same model. However, existing studies are constrained by the following aspects. First, existing research relies on expensive specialized equipment, such as high-quality microphones, which makes real-world applications difficult. Second, recording drone audio in a studio setting is common in research, restricting its practicality in drone delivery scenarios. Third, the resilience of authentication methods against attacks requires validation. Instead, this work proposes a pioneering system *DronAudioID* for lightweight acoustic fingerprint-based drone authentication for secure drone delivery.

III. System and Threat Models

In this section, we illustrate the drone-based logistics transportation process and the threat model of drone delivery.

A. System Model

Logistics companies can utilize drones to offer low-altitude delivery services, enabling contactless and efficient logistics distribution. A sender can place orders through a designated platform, such as a mobile APP provided by the logistics company. The sender specifies the delivery details, including the pickup location, destination address, and some information about the package. Upon receiving the order, the logistics company dispatches a drone to the designated pickup location. The drone, equipped with Global Positioning System (GPS) and navigation systems, can automatically navigate to the pickup location. At the location, the drone lands or hovers at a designated spot where the sender can load the package onto the drone. After that, the drone takes off and follows a predetermined flight path to the destination. Upon arrival at the destination, the receiver receives a notification, indicating that the package has arrived. The receiver can then retrieve the package from the drone at a designated delivery spot.



(a) Impersonation attack

(b) Man-in-the-middle attack

Fig. 2. Threat model of drone delivery.

B. Threat Model

When a sender places an order, it includes detailed delivery information such as sender and receiver addresses, as well as basic package details. A malicious adversary can potentially intercept this information through network attacks. After that, the adversary can launch two kinds of attacks, i.e., impersonation and man-in-the-middle attacks. In the common scenario, it is rare for multiple drones to pick up packages at the same time, typically, the packages are loaded onto one drone before loading the next. Even when multiple drones are present, strict protocols ensure that a safe flying distance is maintained. Therefore, we focus primarily on single-drone scenarios. The threat model is described in Fig. 2.

1) Impersonation Attack: During the shipping phase, the adversary dispatches illegal drones to the pickup location based on intercepted information. Illegal drones can impersonate the legal ones to steal packages. When the sender observes a drone arriving to pick up the packages, it is challenging to determine its legitimacy based on appearance solely, especially if the adversary utilizes the same drone model. Under such an attack, the package is probably loaded onto the illegal drone by the sender himself/herself, thus causing the package loss.

2) Man-in-the-Middle Attack: Even if the drone is legal and transports the packages to the designated receiver, it is susceptible to the man-in-the-middle attack, such attacks could result in the desired packages being substituted during delivery. Since the adversary can obtain some package information, he/she is able to substitute the package during drone delivery. By manipulating the navigation system, the adversary can direct the drone to a predetermined location and swap the

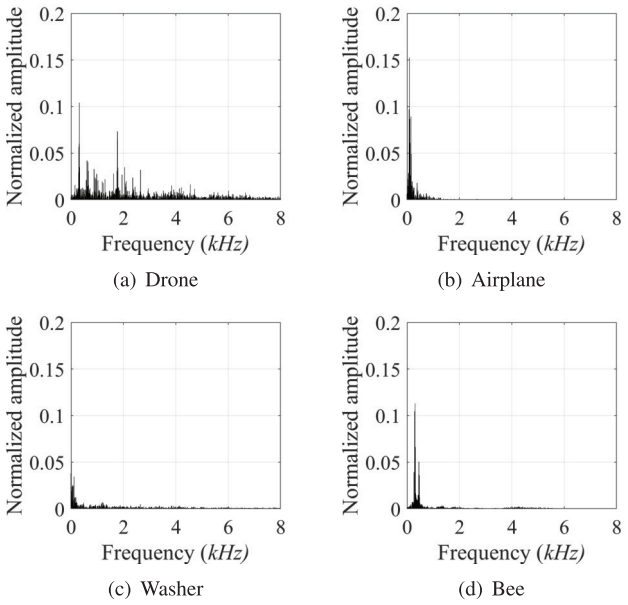


Fig. 3. Spectra of four kinds of audios.

original package. After the substitution, the adversary can restore the address information, letting the drone proceed to the intended recipient. Unlike direct theft of packages, when an attacker substitutes packages during delivery, they can shift responsibility for the incorrect package onto the sender. This action may lead the receiver to believe that the sender failed to send the correct package.

Additionally, there is a scenario where an attacker places an extra object on the drone, intentionally causing the receiver to reject the package, thus resulting in a failed delivery. It is, however, reasonable for the receiver to refuse the package if there is a mismatch between the dispatched and received items, as the additional object could be an unknown and potentially hazardous item. Therefore, the man-in-the-middle attack model defined in this paper focuses primarily on replacing the package to mislead the receiver into believing that the dispatched and received packages are identical.

IV. Drone Audio Characteristics

To implement an acoustic-based drone authentication system and defend against attacks, we first investigate the characteristics of drone audio and explore its distinct fingerprints, serving as the basis of our proposed authentication method.

A. Characteristics of Drone Audio

According to the principle of aerodynamics, a drone flies by pushing the air close to the propellers downwards, as the motor drives the propeller at very high speeds. The primary sources of the drone audio are its rotating propellers and running motor. The spectra of the drone audio and several common audio samples that sound like drone audio are illustrated in Fig. 3. Compared to drone audio, the frequencies of airplane, washer, and bee audio signals are lower. Additionally, the harmonic components of these three kinds of audio signals

TABLE I
Components Differences of Different Drones in the Same Model

Components of drone	Variability factors
Motor bearings	Manufacturing tolerances, wear and tear
Rotor and stator	Design and materials density
Coil windings	Winding density and quality
Motor heat sink	Heat dissipation efficiency
Propeller material and shape	Materials density, Shape differences
Propeller balance	Balance angle
Electronic speed controller	Frequency and response speed of the speed controller
Motor and frame installation	Quality of motor and frame installation
Bolts	Tightness of bolts
Fasteners	Material density and state

are predominantly concentrated below $1kHz$. However, the frequency domain characteristics of the drone are totally different from those of airplane, washer, and bee. Drone audio primarily comprises a fundamental frequency and its corresponding higher-order harmonics. These harmonics exhibit a regular distribution within the range of $0 \sim 8kHz$. The frequencies of these harmonics approximately represent integer multiples of the fundamental frequency.

The fundamental frequency of the drone audio is determined by the KV rating of their stepper motors. The KV rating represents the ratio between the motor's Revolutions Per Minute (RPM) when unloaded and the voltage supplied by the battery. By multiplying the KV rating with the drone battery voltage, we obtain an RPM value which, when divided by 60, yields the fundamental frequency in *Hz*. Different models of drones have different KV ratings for the stepper motors that alter the fundamental frequency and its harmonics.

B. Frequency Characteristics of Drones of the Same Model

The drone audio during flying is mainly generated by propeller and motor vibrations. Although drones of the same model use identical types of engines and rotors, each drone has a unique acoustic fingerprint due to slight variations in hardware manufacturing. The components contributing to this diversity in drone audio are outlined in Table I. For each factor influencing these differences, there are multiple possible variations. By multiplying the number of variations for each factor, it becomes possible to distinguish between thousands of individual drones based on their unique audio profiles.

We recorded the audio of two *DJI Mini*2 drones hovering in identical environmental conditions, whose corresponding acoustic spectra are presented in Fig. 4. It can be observed that the fundamental frequencies of both drone audios are around $320Hz$. This is because the fundamental frequency is determined by the KV ratings and the drone battery voltage. Since the experimental setting for the two drones is controlled, the fundamental frequency of the audio from these two drones is generally similar, with slight variations in their amplitudes only. On the other hand, we can observe significant differences in amplitudes and frequencies in the higher-order harmonic components, which is induced by the imperfection of hard-

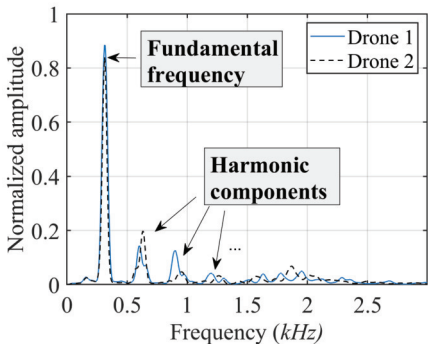


Fig. 4. Spectra of two drones in the same model.

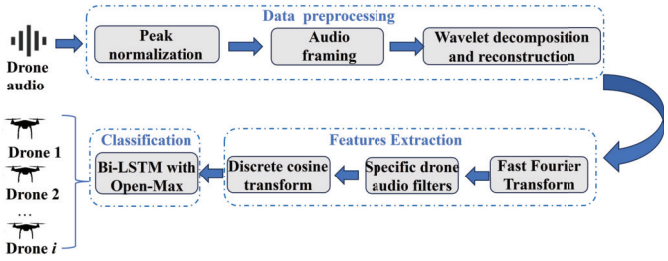


Fig. 5. Signal processing workflow of *DroneAudioID*.

ware manufacturing. Therefore, these unique characteristics demonstrate our intuition and can be utilized to realize the authentication of the different drones of the same model.

V. DroneAudioID Design

In this section, we present the architecture and design details of drone authentication *DroneAudioID*. Fig. 5 illustrates the workflow of *DroneAudioID*. The system comprises three fundamental phases: data preprocessing, feature extraction, and classification. In the data preprocessing phase, we initiate by applying peak normalization to standardize the amplitude levels across different recording signals. Subsequently, each recorded audio file undergoes segmentation into smaller frames to facilitate subsequent phases. Finally, to mitigate the impact of noise on authentication performance, we employ wavelet decomposition and reconstruction. Moving to the feature extraction phase, the audio signal undergoes transformation into the frequency domain via Fast Fourier Transformation (FFT). Following this, we design a specialized filter bank termed Drone Audio Classification Filters (DACF), meticulously crafted to match the spectral characteristics of drone audio, thereby enabling precise feature extraction. Then the Discrete Cosine Transform (DCT) is utilized to derive the feature vector. In the classification phases, particularly for open-set classification, we propose a Bi-LSTM model integrated with the Open-Max algorithm to ensure robust classification, even in scenarios where class labels might not be exhaustive.

A. Data Preprocessing

1) Peak Normalization: Before extracting features from the audio, we should properly pre-process the original audio. First, we perform peak normalization to equalize the amplitude levels of different recording signals. Even in an environment with minimal noise, the mobile device capturing audio signals for the same drone can vary due to minor operational differences, which may cause variations in amplitude levels. Through peak normalization, amplitude dependency in the audio signals can be removed. The mathematical formulation of peak normalization is presented as follows:

$$x'(t) = \frac{x(t) - \min(x(t))}{\max(x(t)) - \min(x(t))}, \tag{1}$$

where $x(t)$ represents the recorded drone audio signal.

2) Audio Framing: The subsequent step involves segmenting each audio file into smaller frames to facilitate feature extraction. Previous studies commonly employed frame durations ranging in $[20 \sim 1,000]ms$. However, the findings [48], [49], [50] show that the audio features extracted from a very short frame length of $20ms$ may not be sufficient for distinguishing drones. Meanwhile, employing a longer frame length results in a reduced size of the extracted feature data, which potentially compromises the performance of classification algorithms. Thus, we choose a frame length of $100ms$, and each frame has a 50% overlap with adjacent frames by default.

3) Denoise: The recorded audio signals comprise various noise components, including environmental noise, and instrument noise from the mobile device. So it is essential to denoise the recorded drone audio. The wavelet transform has a favorable impact on processing audio signals for noise reduction. Therefore, we utilize wavelet transform for signal processing to effectively mitigate noise interference.

By selecting appropriate wavelet basis functions and decomposition levels, the recorded drone audio is decomposed using wavelet analysis. At this point, the signal is mapped onto the wavelet domain. Here it is decomposed into different scales based on wavelet coefficients in spatial domains. The wavelet transform of the drone audio signal can be expressed as:

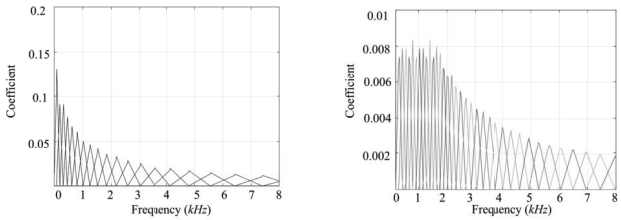
$$x_{wt}(t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x'(t) \cdot \overline{\varphi}\left(\frac{t-b}{a}\right) dt, \tag{2}$$

where a denotes the scale parameter used to control the frequency of the wavelet, b denotes the position parameter used to control the position of the wavelet in time, and $\overline{\varphi}$ denotes the complex conjugate of the wavelet. In this work, we choose a $db8$ wavelet to decompose the signals in 5 layers.

After performing a wavelet transform on the recording audio, the noise-related wavelet coefficients predominantly reside in the high-frequency domain and possess relatively smaller amplitudes. We employ a soft threshold method to eliminate wavelet coefficients with low amplitudes and then use wavelet reconstruction to obtain the denoised audio signals.

B. Feature Extraction

1) FFT: After denoising the audio signals, feature parameters capturing the distinctive characteristics of the target are extracted from the denoised signal and input as feature vectors



(a) MFCC with 26 Mel filters. (b) DACF with 40 specific filters.

Fig. 6. Filters of MFCC and DACF.

into the classifier. During drone hovering, the frequency components of the audio signal remain relatively stable over time. Consequently, FFT is used for frequency domain analysis. The mathematical formulation for FFT processing is as follows:

$$X_k(t) = \sum_{n=0}^{N-1} x_{wt}(n) e^{-j2\pi \frac{k}{N}}, \tag{3}$$

where N represents the number of sampling points, $0 \le k \le N-1$. We set $N = 1323$ in this paper.

Based on the analysis of drone audio characteristics presented in Section III, we observe that complete feature extraction from the entire frequency domain is not essential. This is attributed to the limited energy in the high-frequency domain, rendering it insufficient for drone sound analysis. We conducted an assessment of the average energy distribution across all collected audio samples. It is revealing that nearly 95% of the energy is concentrated within the frequency range of $0 \sim 8kHz$. Hence, we focus on extracting pertinent features within this specific range, which aligns with established practices in drone detection and classification research.

2) Specific Filter Bank Design: After the implementation of FFT, the subsequent step typically involves the conversion of spectral coefficients through MFCC, which is a commonly employed audio feature extraction. MFCC is designed based on the human auditory system. The Mel filterbank simulates how the human ear perceives the frequency of audio signals, while the logarithmic operation mimics the perception of sound intensity by the human ear. Fig. 6(a) shows an MFCC filter bank containing 26 Mel filters. The filter bank consists of more filters in the low-frequency region, while fewer filters are present in the high-frequency region, aiming to capture the nuances of auditory perception. However, in order to distinguish the drones in the same model, it is imperative to design specific filter banks that consistently align with the distinctive characteristics of drone audio. The biggest difference among multiple drones in the same model is the ones in amplitude and frequency of their harmonic components.

Instead of employing MFCC, we design a specific triangle filters named DACF to precisely extract the frequency characteristics of drone audio. According to the study [40], for the majority of commercial drones, their fundamental frequency typically falls within the range of $150 \sim 400Hz$, contingent upon their motor specifications. So we set the center frequency of the first filter as $cenf = 200HZ$.

The harmonics of drones manifest at equidistant intervals within the frequency range of $0 \sim 2kHz$, and these harmonics serve as crucial features for distinguishing drones in the same model. Based on this knowledge, we devised triangular filters with identical bandwidth and amplitude values spanning the $0 \sim 2kHz$ range. Moreover, above $2kHz$, the difference in frequencies and amplitudes in harmonic components between different drones becomes relatively small. Thus we design filters that vary non-linearly with increasing frequency in the range of $2 \sim 8kHz$.

Therefore, our specifically designed filter bank $H_m(k)$ which can be expressed by the following formula:

$$H_m(k) = \begin{cases} 0, & k < f(m-1), k > f(m+1) \\ \dfrac{g(k) - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \le k < f(m) \\ \dfrac{f(m+1) - g(k)}{f(m+1) - f(m)}, & f(m) \le k < f(m+1), \end{cases} \tag{4}$$

where

$$f(m) = \begin{cases} cenf + 1.2 \cdot \dfrac{cenf}{2} \cdot (i-1), & 1 \le i \le 13 \\ f(i-1) \cdot 1.071, & 14 \le i \le m+2, \end{cases} \tag{5}$$

$$g(n) = \frac{n-1}{N} \cdot f_s, \quad 1 \le n \le N. \tag{6}$$

The sampling frequency of drone audio $f_s = 44.1kHz$, and m represents the number of filters. We set $m = 40$ to cover the frequency range of $0 \sim 8\ kHz$. The values of the coefficients in Eq. (5) are determined based on prior experience and insights from relevant literature on triangular filter design [51], [52], [53]. So the DACF which is tailored to drone audio characteristics that contain 40 filters is shown as Fig. 6(b).

3) Feature Vector: To obtain the feature vector for classification, we first calculate the energy of signals in each frame after passing through the filter bank, i.e.,

$$S(i,m) = \sum_{k=0}^{N-1} |X(i,k)|^2 H_m(k), \tag{7}$$

where the variable i represents the number of frames. Then, the DCT is subsequently applied to $S(i,m)$ to perform a cepstral operation. The extraction of cepstral coefficients is:

$$MFC(w,n) = \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} \log\left[S(i,m)\right] \cos\left(\frac{\pi i(2m-1)}{2M}\right), \tag{8}$$

where $1 \le w \le W$, W represents the number of the cepstral coefficients. Finally, we incorporate 40 cepstral coefficients and the energy of each frame into the feature vector, resulting in a total of 41 dimensions as input for classification.

C. Classifier for Open-Set Scenario

1) Bi-LSTM: The neural network models such as LSTM are commonly employed in previous research for classifying drones. LSTM is capable of capturing and retaining long-term dependencies, overcoming the vanishing gradient problem encountered in traditional recurrent neural networks. This enables LSTM to better capture long-term temporal dependencies when processing sequential signals. Compared to LSTM,

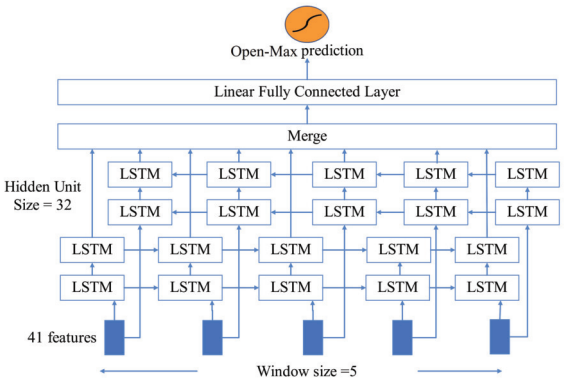


Fig. 7. Bi-LSTM model with Open-Max.

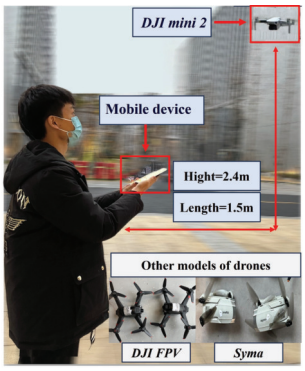


Fig. 8. Outdoor experimental situation.

the Bi-LSTM network enables the analysis of sequential signals in both past and future information. This enables a more comprehensive capture of contextual information. Thus, we employed Bi-LSTM as our classifier due to its promising performance in audio signal classification.

2) Open-Max: Open-Max is an algorithm designed for open-set recognition tasks, aiming to effectively handle samples from unknown classes. In traditional recognition tasks, models are typically trained to identify known classes and assume that all unknown samples belong to one of these known classes. However, in open-world scenarios, models must be able to identify samples from known classes while also detecting and rejecting samples from unknown classes. In practical scenarios, malicious third parties frequently utilize unregistered drones for attacks. Therefore, we developed the Bi-LSTM with Open-Max to enhance the capability of classifying unregistered drones.

First, closed-set training is performed on the Bi-LSTM model using labeled data to learn the features and patterns. During this process, Activation Vectors (AVs) are collected from the trained Bi-LSTM model. These AVs are then passed to the Open-Max layer for analysis, which computes probability scores for each class, including known classes and an unknown class. Finally, a predetermined probability threshold is set to determine whether a sample belongs to a known class or an unknown class. In the classification prediction phase, for each input sample, the model examines the probability values in the Open-Max output vector. If any probability value exceeds the set threshold, the model classifies the sample into the class with the highest probability value. If all probability values are below the threshold, the classifies the sample as the unknown class.

The architecture of the Bi-LSTM with Open-Max is illustrated in Fig. 7. The classification model comprises 2 stacked forward LSTM layers and 2 stacked backward LSTM layers. Each hidden unit possesses a dimensionality of 32. The model is trained for a total of 500 epochs employing the Adam optimizer with a learning rate set to 0.0001. Throughout each epoch, model validation is conducted using the dedicated validation dataset, and the model achieving the lowest loss is saved accordingly. At the classification prediction phase, we average the predictions of multiple consecutive windows and select the class having the highest predicted probability. We determine the optimal number of consecutive windows is 5 through experiments, and the experimental results are presented in Section VI.

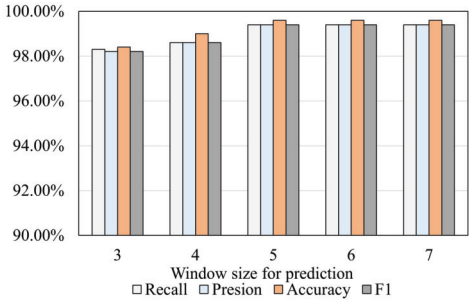
VI. Experimental Evaluations

A. Experimental Setup

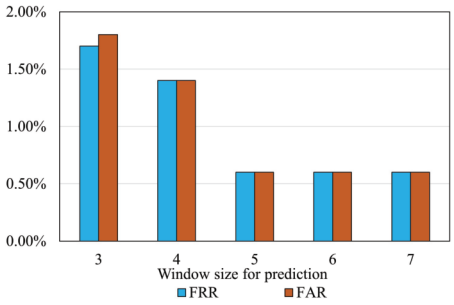
We recorded the drone audio during the hovering phase in an outdoor environment. During the recording process, we selected relatively quiet outdoor locations and avoided recording in places with heavy traffic or high environmental noise pollution. The drone hovering phase often involves steady-state flight with minimal changes in engine pitch or throttle, allowing for clearer authentication of drone audio. The experimenter held the mobile device to record the audio of the drones, which were positioned 2.4m high and 1.5m away from the experimenter. This setup aims to create a more realistic simulation of scenarios involving package shipping and receiving. We separately collected the audios of eight *DJImini*2 drones and we ensured that each drone maintained the same distance from the experimenter. To minimize bias caused by environmental conditions such as weather or temperature, we conducted one experiment each day, recording 3 minutes of audio from each drone hovering, resulting in a total of 10 days of data collection. We used 15% of the total data for prediction, 15% for validation, and 70% for training. Fig. 8 shows the experiment environment.

We use MATLAB's Audio Toolbox to preprocess the collected audio data (such as feature extraction). Then, we use PyTorch's neural network toolkit to train a classification model. The trained PyTorch model is converted to a format suitable for mobile devices via ONNX. On the Android platform, the model is converted to TensorFlow Lite format, and we use the TensorFlow Lite API to load the model. On the iOS platform, we convert the trained model to Core ML format (.mlmodel) and load and run the model using the Core ML framework.

Metrics: We use the following metrics for evaluations. The accuracy is $Accuracy = (TP+TN)/(TP+TN+FP+FN)$, the precision is $Precision = TP/(TP+FP)$, the recall is $Recall = TP\ (TP+FN)$, the F1 score is $\mathrm{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$,



(a)



(b)

Fig. 9. Performance of authentication with different consecutive window sizes.

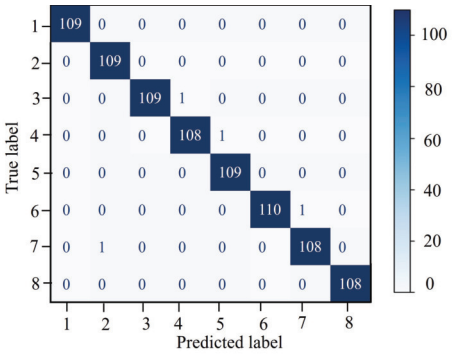
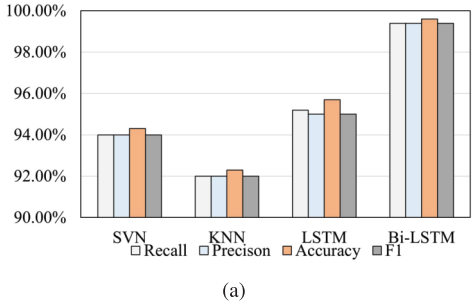


Fig. 10. Confusion matrix of the classification model.

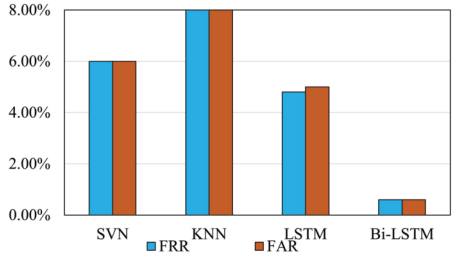
the False Acceptance Rate (FAR) is $FAR = FP/(FP + TN)$, and the False Rejection Rate (FRR) is $FRR = FN/(FN + TP)$, where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

B. Overall Performance

To classify the eight *DJI mini*2 drones, we employ Bi-LSTM model with Open-Max described in Section V-C. Considering that the classification performance can be influenced by the number of consecutive windows involved in decision-making, we first choose the suitable window size for the classification model. Fig. 9 shows the performance metrics under different window sizes. We can see that all the evaluation metrics exhibit an increasing trend with the enlargement of window sizes from 3 to 5. However, beyond a window size of 5, enlarging the window further to 7 fails to yield any noticeable improvement. Instead, it leads to unnecessary resource consumption. Hence, we maintain a consecutive window size of 5 for classification prediction. The corresponding confusion matrix is illustrated in Fig. 10.

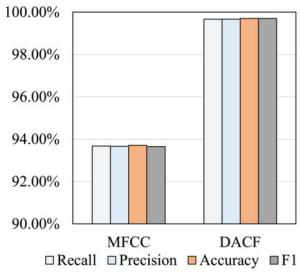


(a)

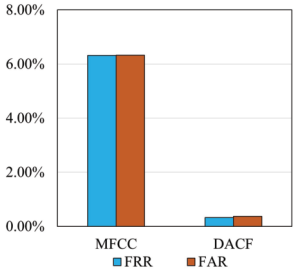


(b)

Fig. 11. Performance of four kinds of classifiers in authentication.



(a)



(b)

Fig. 12. Performance of two feature extraction schemes.

For each authentication, *DroneAudioID* processes the recorded audio files by dividing them into $0.1s$ frames, and the classification result is derived from the outcomes from 5 consecutive time windows. So the user only needs to ensure the recording duration exceeds $0.5s$. After sending the recorded audio to the system, it takes $0.15s$ to get the classification result.

Then, we conducted a comparative analysis of the performance of commonly used classification methods in this field, namely SVM, K-Nearest Neighbors (KNN), LSTM, and Bi-LSTM. Fig. 11 shows the performance of these four methods. We can see that SVM and KNN achieve the *accuracy* of 94.2% and 92.3%, respectively, while both LSTM and Bi-LSTM exhibit *accuracy* exceeding 95%. Notably, the Bi-LSTM model proposed in this paper achieves the lowest *FRR* and *FAR*, alongside an exceptional *accuracy* of 99.6%.

We further conducted a comparison of authentication performance between the MFCC method and the DACF method. Experimental results are illustrated in Fig. 12. Based on MFCC method, the system can only achieve *accuracy* of 93.8%, *precision* of 93.7%, *recall* of 93.7% and $F1$ of 93.7%. *FAR*

TABLE II
COMPARISON OF DIFFERENT METHODS

Method	[47]	[40]	Our method
Environment	Indoor	Outdoor	Outdoor
Device	Professional microphone, recording studio	Professional microphone	Mobile phone
Cost	$6000+	$400+	$200+
Results	Individual drone	Drone type	Individual drone
Recall	96.2%	95.2%	99.4%
Precision	96.3%	95.0%	99.4%
Accuracy	96.2%	95.7%	99.6%
$F1$	96.3%	95.0%	99.4%
FRR	3.8%	4.8%	0.6%
FAR	3.7%	5.0%	0.6%

and *FRR* both are 6.3%. While all the evaluation metrics of the DACF-based method proposed in this paper outperform those of the MFCC-based method.

Additionally, we provide a comparison with state-of-the-art methods, emphasizing key differences and improvements, as outlined in Table II. We can see that *DroneAudioID* outperforms other methods in authentication performance while requiring only a mobile device for audio recording. Through the data preprocessing, feature extraction, and classification models described in Section V, users can perform authentication using a widely available, portable mobile phone with built-in recording capabilities. In contrast, other methods typically rely on specialized recording equipment or controlled environments. As a result, *DroneAudioID* offers a more convenient and cost-effective authentication process, highlighting its lightweight design and practical advantages.

C. Security Analysis

1) Impersonation Attacks: After the sender places an order through the APP, the logistics company dispatches a registered drone to pick up the package. Meanwhile, an attacker gains access to a drone of the same model and steers the illegal drone towards the designated pickup location for stealing the package. In such a situation, the attacker attempts to replay the audio of a legal drone to deceive our authentication system by impersonating a legal drone. The recordings can be surreptitiously obtained by attackers during previous deliveries made by legal drones.

Due to the characteristics of drone audio, *DroneAduioID* can differentiate an impersonating drone from a legal one. Fig. 13 shows the spectra of impersonating and legal drones. We can see that a notable distinction arises between the frequency characteristics of the impersonating drone and the legal one. This disparity primarily stems from two factors. First, When the impersonating drone uses a player to replay the audio of a legal one, the software and hardware design of the player automatically perform some signal processing technologies on the audio signals, like filtering, thereby changing the characteristics of the audio. Second, while recording, the mobile device also captures drone audio emitted by the impersonator itself. The final audio recorded for authentication primarily consists of a combination of these two kinds of audio. These are the reasons why the *DroneAudioID* can differentiate an impersonating drone from a legal one.

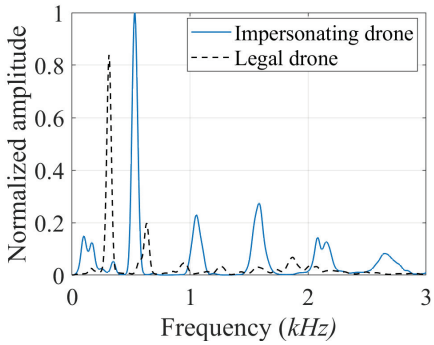


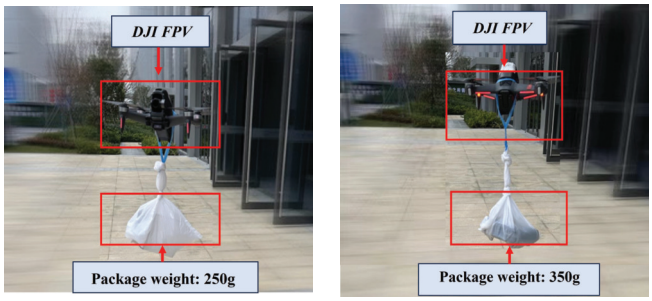
Fig. 13. Spectra of impersonating drone and legal drone.

To illustrate this attack scenario, we first recorded the audio of one of the legal *DJI mini*2 drones and then employed a *MarshallEmbertonII* Bluetooth player to replay the recording while the other drone was hovering. The experimental results show that *DroneAduioID* can achieve *accuracy* of 98.5%, *precision* of 98.2%, *recall* of 98.2% and $F1$ of 98.2%. Both the *FRR* and *FAR* are 1.8%. This shows that our system can effectively resist impersonation attacks.

2) Man-in-the-Middle Attacks: After the legal drone picks up the package, the package will be delivered to the designated receiver. However, when transporting valuable packages, the malicious adversary may substitute the original package during transportation. The package substitution leads to a deviation in weight. Typically, it is challenging for an adversary to find a substitute package with exactly the same weight. Even if the attacker manages to match the weight, differences in the distribution of the center of mass between items would persist. This would force the drone's engines and rotors to operate at different frequencies to maintain flight stability, resulting in distinct audio signatures. Moreover, while theoretically possible, the cost for the attacker to replicate these complex features, such as precisely adjusting the center of mass or designing a custom-shaped package, would be prohibitively high. Increasing the attack cost is a fundamental strategy in security design. Therefore, although such an attack could be theoretically feasible, the practical costs and complexities make it highly unlikely to be a real-world threat.

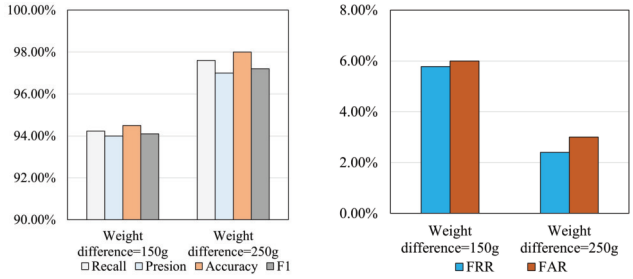
As the different package weights carried by the drone lead to various levels of audio intensity, *DroneAduioID* can utilize this to authenticate whether the package is substituted.

To demonstrate it, we first recorded the audio of a drone carrying a package weighing 100g. Similarly, we conducted one experiment each day, recording 3 minutes of audio from each drone hovering, resulting in a total of 10 days of data collection. The entire dataset was then used for model training. Additionally, under the same experimental conditions, we recorded audio from the same drone carrying packages weighing 250g and 350g, respectively. So the differences in package weights are 150g and 250g. The scenarios of drone carrying packages are depicted in Fig. 14.



(a) Weight difference=150g (b) Weight difference=250g

Fig. 14. Drone equipped with package of varying weights.



(a) (b)

Fig. 15. Performance of the system under different package weight differences.

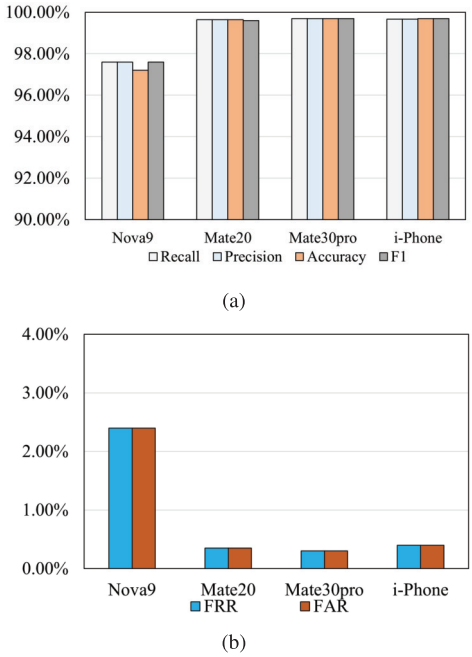
TABLE III
NUMBER OF MEASUREMENTS FOR EACH MODEL OF MOBILE DEVICE AND DRONE

The number of measurements	$DJI\ mini2$	$Syma$	$DJI\ FPV$
$Huawei\ Mate30Pro$	10	10	10
$Huawei\ Mate20$	10	-	-
$Nova9$	10	-	-
$iPhone14Pro$	10	-	-

The authentication performance is shown in Fig. 15. When the weight difference is 150g, the *DroneAduioID* can achieve *accuracy* of 94.5%, *precision* of 94.0%, *recall* of 94.2% and $F1$ of 94.1%. The *FRR* is 5.8% and the *FAR* is 6.0%. When the weight difference is 250g, *accuracy* of 98.0%, *precision* of 97.0%, *recall* of 97.6% and $F1$ of 97.3% can be achieved. The *FRR* is 2.4% and the *FAR* is 3.0%. These results show that the larger the weight difference is, the better performance we can get.

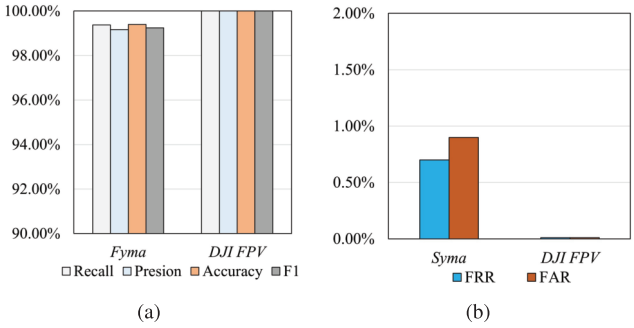
D. Impact of Different Models of Mobile Devices and Drones

To verify the universality of *DroneAudioID*, we conducted tests on different models of drones and mobile devices. The combination of test devices and the drone models to be tested is shown in Table III. First, to discuss the impact of different types of mobile devices on the authentication performance, we separately recorded the drone audio of *DJImini*2 by using four types of mobile phones, namely: *HuaweiMate*20, *HuaweiMate*30*Pro*, *Nova*9, and *i-Phone*14*Pro*. We trained the classification model using data collected from all four types of mobile phones, ensuring a comprehensive dataset that reflects the variations across devices.



(a) (b)

Fig. 16. Performance of the system authenticating different models of mobile phones.



(a) (b)

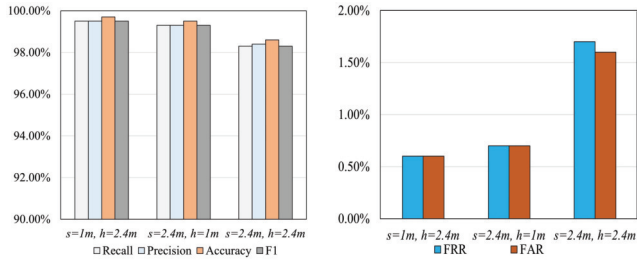
Fig. 17. Performance of the system authenticating different models of drones.

As the authentication performance shown in Fig. 16(a), by using *HuaweiMate*20, *HuaweiMate*30*Pro*, and *i-Phone*14*Pro*, the *DroneAudioID* can achieve similar authentication performance in which all the evaluation metrics pass 99.5%. By using *Nova*9, we can achieve the evaluation metrics around 97.2%. The *FRR* and *FAR* of authentication are shown in Fig. 16(b).

Second, we discuss the impact of different types of drones on authentication performance. We used the *HuaweiMate*30*Pro* to record the two additional models respectively, *Syma* drones and *DJIFPV* drones. We implemented *DroneAudioID* on two *Syma* drones, achieving *accuracy* of 99.4%, *precision* of 99.1%, *recall* of 99.3% and $F1$ of 99.2%. The *FRR* is 0.7% and the *FAR* is 0.9%. All the evaluation metrics reach 100% when authenticating the two *DJIFPV* drones. The authentication performance is shown in Fig. 17.

E. Impact of Different Distances

When loading and unloading packages, the variation in distance between the user and the drone can potentially impact



(a) (b)

Fig. 18. Performance of the system with different distances.

the authentication performance. To address this, we introduce two parameters: s representing the horizontal distance between the drone and the user. And h denoting the vertical distance. Consequently, we combine different s and h to assess our system. Considering that the receiver needs to maintain a certain distance from the drone for authentication, yet the distance between them cannot be too far, we set the maximum values of $s = 2.4m$ and $h = 2.4m$. The authentication performance with different distances is presented in Fig. 18. As shown in Fig. 18(a), all the evaluation metrics exceed 99% under both scenarios: where $s = 2.4m$, $h = 1m$, as well as $s = 1m$, $h = 2.4m$. Even in the situation where both s and h are set to $2.4m$, all the evaluation metrics exceed 98%. The *FRR* and *FAR* are shown in Fig. 18(b).

F. Impact of Different Signal-to-Noise Ratio

When recording drone audio in the outdoor environment, different levels of noise may be present. To assess the robustness of *DroneAudioID* in diverse environments, we download the noise samples of three scenarios (i.e., canteen, city, and farm) from YouTube. We incorporate each noise sample into the original drone audio at Signal-to-Noise Ratio (SNR) of $1dB$, $5dB$, and $10dB$ respectively. The performance of the system under different SNRs with three kinds of scenarios is presented in Fig. 19.

As we can see in the Fig. 19. The authentication performance in three different scenarios improves as the SNR increases. Specifically, when SNR = $1dB$, which means the strength of noise and signal is very close, the recording environment is extremely harsh. All evaluation metrics have exceeded 87% in both the canteen and city scenarios. However, the authentication performance is marginally lower in the farm scenario compared to the other two scenarios, which just pass 71%. This is because in the farm scenario, the instantaneous noise such as birds' sounds and insect chirping can significantly influence the characteristics of drone audio. In this situation, users are advised to wait momentarily until the brief intermittent sound ceases before initiating the authentication process. When SNR = $5dB$, all the evaluation metrics exceed 91% in both canteen and city scenarios. Specifically in the farm scenario, the authentication performance has seen a significant improvement, with all evaluation metrics surpassing 87%. When SNR = $10dB$, all the evaluation metrics exceed 93% in three scenarios. In practice, achieving an SNR of $10dB$ during the authentication process is a prevalent and easily attainable outcome. The *FRR* and *FAR* of the authentication are summarized in Tabel IV.

TABLE IV
FRR AND *FAR* UNDER DIFFERENT SCENARIOS

	Carteen		
	SNR=$1dB$	SNR=$5dB$	SNR=$10dB$
FRR	12.8%	8.5%	2.5%
FAR	8.3%	7.8%	2.1%
	City		
	SNR=$1dB$	SNR=$5dB$	SNR=$10dB$
FRR	11.4%	8.7%	4.9%
FAR	8.9%	7.1%	4.1%
	Farm		
	SNR=$1dB$	SNR=$5dB$	SNR=$10dB$
FRR	29.0%	12.3%	7.8%
FAR	26.7%	8.9%	7.0%

VII. Discussion

A. Complex Environment

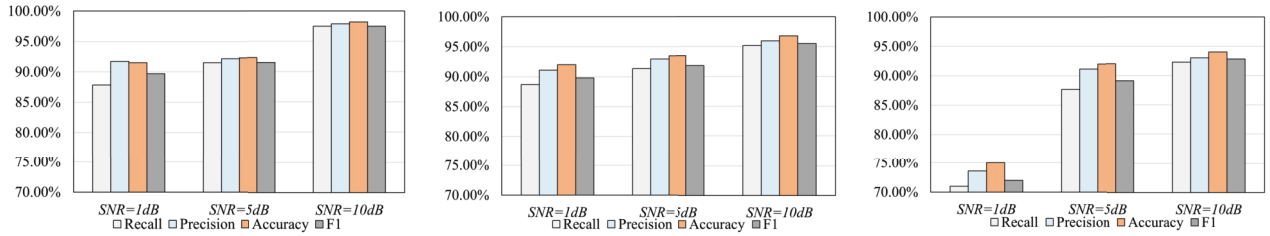
We have assessed the robustness of our authentication system against different types of noise at varying *SNR*s. However, the dynamic and unpredictable nature of the farm environment presents unique challenges, particularly due to sudden bursts of noise like bird calls and insects chirping during authentication. To address this issue, in our future study, we plan to investigate denoising techniques such as adaptive filters and deep learning models to mitigate the impact of this burst noise on authentication performance. Adaptive filters offer real-time noise mitigation by dynamically adjusting to changing noise profiles, while deep learning models like CNNs and Recurrent Neural Networks (RNNs) have shown promise in extracting meaningful signal information from noisy observations.

B. Drone Swarm Scenario

In this paper, we have developed and tested our drone authentication system within the framework of a single drone. However, as we broaden our scope to address package delivery by drone swarms, we encounter a new set of complexities. In this scenario, the overlapping audio signals produced by multiple drones introduce significant interference, posing a substantial obstacle to authentication. To overcome this challenge, we're exploring the integration of Independent Component Analysis (ICA) technology alongside blind source separation techniques. By harnessing the power of ICA, we aim to effectively segregate the mixed and intricate audio generated by the drone swarm. This approach enables us to disentangle the overlapping signals and isolate the drone audio signals from the ambient noise and interference. This enhances the reliability and accuracy of our authentication system in multi-drone scenarios.

C. Wear and Assemble

In our experimental observations, we identified an intriguing phenomenon associated with the wear and tear drones undergo during extended flight hours. As components such as propellers and motors inevitably degrade, they subtly alter



(a) Canteen (b) City (c) Farm

Fig. 19. Performance of the system under different SNRs.

the drone's acoustic fingerprint. While these changes in audio characteristics are relatively minor, they have a noticeable impact on the authentication process. To further investigate this, we conducted additional experiments. Specifically, we collected audio data from the registered drone after 40 hours of fight. To simplify the analysis, we focused solely on cases where the drone showed no visible damage and applied the trained model for authentication. By using *DroneAudioID*, we can get *accuracy* of 96.5%, *precision* of 96.6%, *recall* of 96.6% and $F1$ of 96.6%. The *FAR* and *FRR* both are 3.4%. Moreover, when the drone sustains visible damage or undergoes component replacements, its acoustic fingerprint shifts significantly. We are currently developing a dynamic re-registration protocol to account for such hardware changes, ensuring the system's accuracy and security are maintained.

VIII. Conclusion

In this paper, we propose *DroneAudioID*, which is a lightweight acoustic fingerprint-based drone authentication system for secure drone delivery. We employ acoustic fingerprints to authenticate different drones of the same model based on differences in fundamental frequency and harmonic components of the drone audios. First, we apply wavelet transform to remove high-frequency noise during data preprocessing. Then, a specialized filter bank named DACF is designed for drone audio feature extraction. Subsequently, we construct a Bi-LSTM with Open-Max for open-set classification. The experimental results demonstrate that *DroneAudioID* can achieve an accuracy of 99.6% to distinguish the different drones of the same model. Also, it can effectively defend against impersonation attacks based on drone audio replay and package substitution attacks, with accuracies of 98.5% and 98.0%, respectively.

References

[1] B. Shahzaad, A. Bouguettaya, S. Mistry, and A. G. Neiat, "Composing drone-as-a-service (DaaS) for delivery," in *Proc. IEEE Int. Conf. Web Services (ICWS)*, Apr. 2019, pp. 28–32.

[2] Amazon. (2016). *Amazon Prime Air*. [Online]. Available: https://amzn.to/2oFPnmj

[3] E. C. Ferrer, "The blockchain: A new framework for robotic swarm systems," in *Proc. Springer FTC*. Cham, Switzerland: Springer, 2019, pp. 1037–1058.

[4] A. Chapman. (2020). *Drones Find Rising Role in Agriculture*. [Online]. Available: https://www.beefcentral.com/ag-tech/drones-and-automatedvehicles/drones-find-rising-role-in-precision-agriculture/

[5] K. Zetter. (2013). *Gaming Company Certificates Stolen and Used to Attack Activists, Others*. [Online]. Available: https://www.wired.com/2013/04/gaming-company-certs-stolen/

[6] L. Wang et al., "DF-sense: Multi-user acoustic sensing for heartbeat monitoring with dualforming," in *Proc. 21st Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2023, pp. 1–13.

[7] L. Lu et al., "VPad: Virtual writing tablet for laptops leveraging acoustic signals," in *Proc. IEEE 24th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2018, pp. 244–251.

[8] L. Lu et al., "Enable traditional laptops with virtual writing capability leveraging acoustic signals," *Comput. J.*, vol. 64, no. 12, pp. 1814–1831, Dec. 2019.

[9] Y. Wu, F. Li, Y. Xie, Y. Wang, and Z. Yang, "SymListener: Detecting respiratory symptoms via acoustic sensing in driving environments," *ACM Trans. Sensor Netw.*, vol. 19, no. 1, pp. 1–21, Feb. 2023.

[10] Y. Bai, J. Liu, Y. Chen, L. Lu, and J. Yu, "Poster: Inaudible high-throughput communication through acoustic signals," in *Proc. ACM MobiCom*, 2019, pp. 1–3.

[11] Y. Bai, J. Liu, L. Lu, Y. Yang, Y. Chen, and J. Yu, "BatComm: Enabling inaudible acoustic communication with high-throughput for mobile devices," in *Proc. 18th Conf. Embedded Networked Sensor Syst.*, Nov. 2020, pp. 205–217.

[12] T. Sun, Y. Zhao, W. Xie, J. Li, Y. Ma, and J. Zhang, "EyeGesener: Eye gesture listener for smart glasses interaction using acoustic sensing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 8, no. 3, pp. 1–28, Aug. 2024.

[13] L. Lu et al., "LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Jul. 2018, pp. 1466–1474.

[14] L. Lu et al., "Lip reading-based user authentication through acoustic sensing on smartphones," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 447–460, Feb. 2019.

[15] L. Lu, J. Yu, Y. Chen, and Y. Wang, "Vocallock: Sensing vocal tract for passphrase-independent user authentication leveraging acoustic signals on smartphones," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 51:1–51:24, 2020.

[16] L. Lu et al., "KeyListener: Inferring keystrokes on QWERTY keyboard of touch screen through acoustic signals," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2019, pp. 775–783.

[17] J. Yu, L. Lu, Y. Chen, Y. Zhu, and L. Kong, "An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 337–351, Feb. 2021.

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

[19] 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.

[20] H. C. Kumawat, M. Chakraborty, and A. A. B. Raj, "DIAT-RadSATNet—A novel lightweight DCNN architecture for micro-doppler-based small unmanned aerial vehicle (SUAV) targets' detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[21] Y. Sun, S. Abeywickrama, L. Jayasinghe, C. Yuen, J. Chen, and M. Zhang, "Micro-Doppler signature-based detection, classification, and localization of small UAV with long short-term memory neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6285–6300, Aug. 2021.

[22] C. Schüpbach, C. Patry, F. Maasdorp, U. Böniger, and P. Wellig, "Micro-UAV detection using DAB-based passive radar," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1037–1040.

[23] Á. D. de Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. López, "Drone detection with X-band ubiquitous radar," in *Proc. 19th Int. Radar Symp. (IRS)*, Jun. 2018, pp. 1–10.

[24] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged Doppler images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, Jan. 2017.
[25] P. Zhang, G. Li, C. Huo, and H. Yin, "Exploitation of multipath micro-Doppler signatures for drone classification," *IET Radar, Sonar Navigat.*, vol. 14, no. 4, pp. 586–592, Apr. 2020.
[26] J. Park and J.-S. Park, "Classification of small drones using low-uncertainty micro-Doppler signature images and ultra-lightweight convolutional neural network," *IEEE Trans. Image Process.*, vol. 33, pp. 2979–2994, 2024.
[27] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and Bluetooth interference," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 60–76, 2020.
[28] P. Nguyen, H. Truong, M. Ravindranathan, A. Nguyen, R. Han, and T. Vu, "Cost-effective and passive RF-based drone presence detection and characterization," *GetMobile: Mobile Comput. Commun.*, vol. 21, no. 4, pp. 30–34, Feb. 2018.
[29] Z. Li et al., "Reliable digital forensics in the air: Exploring an RF-based drone identification system," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–25, Jul. 2022.
[30] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Micro-UAV detection and classification from RF fingerprints using machine learning techniques," in *Proc. IEEE Aerosp. Conf.*, Mar. 2019, pp. 1–13.
[31] C. J. Swinney and J. C. Woods, "RF detection and classification of unmanned aerial vehicles in environments with wireless interference," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2021, pp. 1494–1498.
[32] F. Mahdavi and R. Rajabi, "Drone detection using convolutional neural networks," in *Proc. 6th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS)*, Dec. 2020, pp. 1–5.
[33] M. W. Ashraf, W. Sultani, and M. Shah, "Dogfight: Detecting drones from drones videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7063–7072.
[34] D. K. Behera and A. B. Raj, "Drone detection and classification using deep learning," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 1012–1016.
[35] Q. Shi and J. Li, "Objects detection of UAV for anti-UAV based on YOLOv4," in *Proc. IEEE 2nd Int. Conf. Civil Aviation Saf. Inf. Technol. (ICCASIT*, Oct. 2020, pp. 1048–1052.
[36] Q. Dong, Y. Liu, and X. Liu, "Drone sound detection system based on feature result-level fusion using deep learning," *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 149–171, Jan. 2023.
[37] İ. Aydın and E. Kızılay, "Development of a new light-weight convolutional neural network for acoustic-based amateur drone detection," *Appl. Acoust.*, vol. 193, May 2022, Art. no. 108773.
[38] A. Bernardini, F. Mangiatordi, E. Pallotti, and L. Capodiferro, "Drone detection by acoustic signature identification," *Electron. Imag.*, vol. 29, pp. 60–64, Jan. 2017.
[39] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine learning inspired sound-based amateur drone detection for public safety applications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2526–2534, Mar. 2019.
[40] H. Kolamunna et al., "DronePrint: Acoustic signatures for open-set drone detection and identification with online data," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 1, pp. 1–31, 2021.
[41] O. A. Ibrahim, S. Sciancalepore, and R. Di Pietro, "Noise2Weight: On detecting payload weight from drones acoustic emissions," *Future Gener. Comput. Syst.*, vol. 134, pp. 319–333, Sep. 2022.
[42] C. Wu and Q. Zeng, "Turning noises to fingerprint-free 'credentials': Secure and usable drone authentication," *IEEE Trans. Mobile Comput.*, vol. 23, no. 10, pp. 10161–10174, Oct. 2024.
[43] C. Wu, X. Li, L. Luo, and Q. Zeng, "G2Auth: Secure mutual authentication for drone delivery without special user-side hardware," in *Proc. 20th Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2022, pp. 84–98.
[44] H. Yang et al., "Wave-for-safe: Multisensor-based mutual authentication for unmanned delivery vehicle services," in *Proc. ACM MobiHoc*, 2023, pp. 230–239.
[45] J. O. Sharp, C. Wu, and Q. Zeng, "Authentication for drone delivery through a novel way of using face biometrics," in *Proc. ACM MobiCom*, Oct. 2022, pp. 609–622.
[46] S. Ramesh, T. Pathier, and J. Han, "SoundUAV: Towards delivery drone authentication via acoustic noise fingerprinting," in *Proc. ACM DroNet*, Seoul, South Korea, 2019, pp. 27–32.
[47] Y. Diao, Y. Zhang, G. Zhao, and M. Khamis, "Drone authentication via acoustic fingerprint," in *Proc. 38th Annu. Comput. Secur. Appl. Conf.*, Austin, TX, USA, Dec. 2022, pp. 658–668.
[48] E. Babaee, N. B. Anuar, A. W. A. Wahab, S. Shamshirband, and A. T. Chronopoulos, "An overview of audio event detection methods from feature extraction to classification," *Appl. Artif. Intell.*, vol. 31, nos. 9–10, pp. 661–714, Nov. 2017.
[49] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669–138682, 2019.
[50] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, "A survey of audio classification using deep learning," *IEEE Access*, vol. 11, pp. 106620–106649, 2023.
[51] S.-M. Lee, S.-H. Fang, J.-W. Hung, and L.-S. Lee, "Improved MFCC feature extraction by PCA-optimized filter-bank for speech recognition," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Aug. 2001, pp. 49–52.
[52] J.-W. Hung, "Optimization of filter-bank to improve the extraction of MFCC features in speech recognition," in *Proc. Int. Symp. Intell. Multimedia, Video Speech Process*, 2004, pp. 675–678.
[53] S. Chakroborty and G. Saha, "Improved text-independent speaker identification using fused MFCC & IMFCC feature sets based on Gaussian filter," *Int. J. Signal Process.*, vol. 5, no. 1, pp. 11–19, 2009.

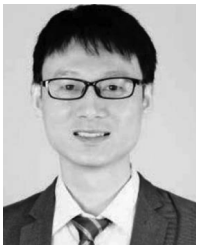
Meng Zhang received the Ph.D. degree from the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. She is currently a Post-Doctoral Researcher with Zhejiang University. Her current research interests include mobile sensing, wireless communication security, the Internet of Things security, AI security, and the IoT applications.

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

Yuhan Wu received the B.E. degree from the College of Computer Science and Technology, Zhejiang University. He is currently pursuing the master's degree with the Polytechnic Institute of Zhejiang University. His current research interests include the Internet of Things security, vehicle security, and AI security.

Zheng Yan is currently pursuing the bachelor's degree with the College of Computer Science and Technology, Zhejiang University. His research interests include the Internet of Things security and vehicle security.

Jiaqi Sun is currently pursuing the bachelor's degree with the College of Computer Science and Technology, Zhejiang University. His research interests include vehicle security and malware detection.



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

Kui Ren (Fellow, IEEE) received the B.Eng. degree in chemical engineering and the M.Eng. degree in materials engineering from Zhejiang University, China, in 1998 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from Worcester Polytechnic Institute, USA, in 2007. He is currently the Dean of the College of Computer Science and Technology, Zhejiang University. He is mainly engaged in research in data security and privacy protection, AI security, and security in intelligent devices and vehicular networks. He is a fellow of AAAS, ACM, and CCF.