Unleashing the Power of Vibration and Acoustic Signals for Fast and Reliable Communication

Zhen Guo
Saint Louis University, USA
zhen.guo.2@slu.edu

Wei Wang
Texas Tech University, USA
wei.wang@ttu.edu

Reza Tourani
Saint Louis University, USA
reza.tourani@slu.edu

***Abstract*—Physical vibration-based communication has recently become an attractive communication paradigm owing to its diverse range of applications across various domains. However, until now, a communication system that supports fast and reliable vibration communication using commodity devices is still missing. In this paper, we present Vib-Sound– the first communication system that brings fast and reliable vibration communication to existing commodity devices, such as smartphones and tablets. The high-level idea of Vib-Sound is to enable simultaneous collaboration between the vibration and corresponding side-channel acoustic signals generated by the motor in a smart device. The core design of Vib-Sound includes *(i)* a novel Interference-Aware Vibration Modulation technique that generates the optimal vibration signals, *(ii)* a Multi-Medium Channel Modeling scheme to conduct vibration and acoustic signal compensation, and *(iii)* a new Collaborative Cross-Channel Demodulation approach for real-time vibration and acoustic signal demodulation utilizing physics-informed cost functions. This ensures the learned demodulation model is aligned with the underlying physical principles and constraints governing vibration and acoustic signal transmission. We conduct extensive real-world experiments under various scenarios and settings. The evaluation results show the effectiveness of our design.**

***Index Terms*—vibration, acoustics, cross-channel, interference-aware modulation, collaborative demodulation, authentication.**

I. Introduction

Physical vibration-based communication has emerged as a versatile paradigm for applications ranging from secure data transfer [1]–[6] and authentication [7]–[9] to material sensing [10] and industrial fault detection [11]. By modulating a sender's vibration motor to encode data, a receiver can demodulate the transmitted bits by sensing physical oscillations.

Despite this potential, existing systems face a trade-off between hardware complexity and throughput. Pioneer works like Ripple I/II [1], [2] and VibroComm [12] achieved high speeds but required specialized piezo transducers or mmWave radar for reliability. Conversely, while commodity-device solutions offer better practicality, they are limited to low-speed regimes (< 40 bps). Consequently, a system capable of achieving high-speed, reliable vibration communication using only unmodified commodity hardware remains an open challenge.

We introduce *Vib-Sound,* the first system that achieves fast and reliable vibration communication using commodity devices. The **high-level idea** of our approach is to enable simultaneous collaboration between the vibration signal and its side-channel acoustic signal generated by the motor. Specifically, the sender (*e.g.,* smartphone) will carefully control the motors for modulation. Subsequently, rather than solely demodulating the vibration signal using the embedded accelerometer, the receiver will also use the microphone to demodulate the corresponding acoustic signal simultaneously.

The design of Vib-Sound using commodity devices is challenging. Specifically, to achieve fast and reliable communication, one of the critical steps is to design a modulation scheme to modulate the vibration signals effectively. Due to the hardware limitations of commodity devices, the sender (*e.g.,* smartphone) cannot control the motor accurately. As a result, the modulated vibration symbols will suffer inevitable distortions, significantly reducing the communication's throughput and reliability. To overcome this challenge, we introduce a novel **Interference-Aware Vibration Modulation (IVM)** technique to *(i)* determine the optimal symbol length for throughput improvement and *(ii)* find the desired modulation scheme to improve the communication reliability. Moreover, vibration signals propagate through surfaces, while acoustic signals travel through the air. Hence, changes in surface conditions and environmental noise can distort the transmitted signals. To overcome this challenge, we introduce a new **Multi-Medium Channel Modeling (MCM)** scheme to model the vibration and acoustic channels for signal compensation.

Vib-Sound also introduces a **Collaborative Cross-Channel Demodulation (CCD)** approach to further improve the throughput and reliability. To be more specific, traditional demodulation approaches [1] require the motor to stop completely before transmitting the next symbol, ensuring demodulation reliability at the cost of reduced throughput. In our approach, CCD uses a novel ensemble *signal sequence and spectrum united model* to analyze and demodulate the vibration signal, spectrum information, and the corresponding side-channel acoustic signal simultaneously and does not require the motor in a sender to stop completely before transmitting the next symbol. Nonetheless, the limited sampling data at the receiver due to accelerometer hardware limitations, vibration signal attenuation, and ambient noise adversely impact demodulation reliability. To overcome this challenge and further enhance demodulation reliability, we formulate the CCD approach as a physics-informed learning-based demodulation. Specifically, we introduce novel physics-informed cost functions customized for the reliable processing

ISBN 978-3-903176-82-9 © 2026 IFIP

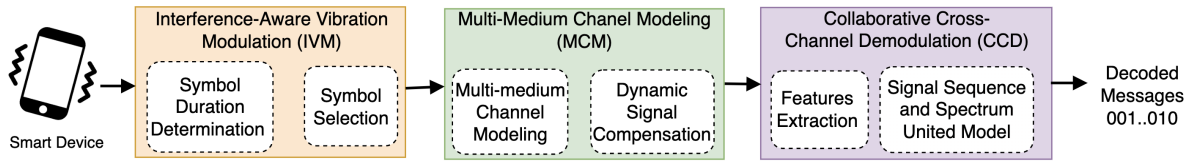

Fig. 1: Vib-Sound consists of three main modules: Interference-Aware Vibration Modulation (IVM), Multi-Medium Channel Modeling (MCM), and Collaborative Cross-Channel Demodulation (CCD).

of received signals by incorporating the underlying physical principles of vibration and acoustic signal transmission.

In summary, the **contributions** of this paper are as follows:

• To the best of our knowledge, Vib-Sound is the first work that achieves fast and dependable vibration communication using commodity devices, *e.g.,* smartphones and tablets. The design principles, challenges, and solutions in Vib-Sound are generic and applicable to commodity devices and future vibration communication systems.

• By introducing novel Interference-Aware Vibration Modulation, Multi-Medium Channel Modelling, and Collaborative Cross-Channel Demodulation, Vib-Sound achieves seamless collaboration between vibration signals and corresponding acoustic signals generated by the motor for fast and reliable vibration communication.

• We implement Vib-Sound on commodity devices and extensively evaluated our design under different settings and real-world scenarios. The evaluation results show Vib-Sound almost doubles the throughput compared to the most recent solution and can conduct reliable communication under extremely noisy environments.

II. Motivating Use Cases

We highlight the necessity for high-speed, commodity-based vibration communication through two key scenarios where specialized hardware is impractical.

Use Case 1: Secure Authentication and Key Exchange. Vibration and acoustic modalities are increasingly used for multi-factor authentication and device pairing [7]. However, a performance gap exists: high-speed schemes (e.g., AM/FM) typically necessitate customized hardware [1], [2], while commodity-compatible On-Off Keying (OOK) suffers from low throughput, limiting its use in complex security protocols. Vib-Sound bridges this gap by enabling high-speed data transfer on standard devices. We validate this via a 20-participant proof-of-concept authentication study (Section V-C), demonstrating its robustness in real-world settings.

Use Case 2: Out-of-Band (OOB) IoT Communication. With nearly 15 billion IoT devices competing for the crowded ISM bands, CSMA-induced contention leads to significant latency and throughput degradation [13], [14]. Vib-Sound provides a reliable OOB signaling channel that operates independently of the electromagnetic spectrum. By offloading control signaling or small data transfers to the vibration-acoustic domain, Vib-Sound mitigates RF interference and preserves bandwidth for primary wireless traffic, enhancing overall network capacity without requiring additional hardware.

III. Design Overview

The design goal of Vib-Sound is to achieve fast and reliable communication using vibration and acoustic signals. To achieve it, as shown in Figure 1, our system mainly consists of three key modules:

• **Interference-Aware Vibration Modulation (IVM).** In Vib-Sound, we introduce the Interference-Aware Vibration Modulation component to modulate the vibration signals effectively. Specifically, to achieve high-throughput communication, we first introduce a ***symbol duration determination*** algorithm to find the optimal length for the vibration signal. Then, to achieve reliable communication, we propose a novel ***vibration symbol selection*** scheme to select the optimal symbols with the largest Hamming distance between each other (§ IV-A).

• **Multi-Medium Chanel Modeling (MCM).** To further improve communication reliability, we introduce Multi-Medium Channel Modeling (MCM) to conduct signal compensation at the receiver side. MCM solves the fundamental problem of how to model the multi-medium channel (*i.e.,* solid surface and air) using distorted vibration and corresponding acoustic signals generated by the motor. According to the channel model, the receiver can conduct dynamic signal compensation for reliable communication (§ IV-B).

• **Collaborative Cross-Channel Demodulation (CCD).** In Vib-Sound, we introduce a Collaborative CCD scheme to simultaneously demodulate the vibration and acoustic signals. Considering the inherent reliability-throughput tradeoff in vibration and acoustic communication, we developed physics-informed loss functions by incorporating principles on signal attenuation and multi-path effects, aiming to enhance reliability while preserving high throughput capabilities. Specifically, CCD first extracts the unique features from the vibration and corresponding acoustic signals in the time domain. Then, it leverages spectrum information of the vibration signals to improve the demodulation granularity. At last, all the features are seamlessly fed into a novel lightweight ***signal sequence and spectrum united model*** to demodulate the signals (§ IV-C).

IV. Detailed Design

A. Interference-Aware Vibration Modulation

The objective of IVM is to encode data into robust vibration-silence sequences. We define the minimal detectable vibration duration as an *atom* (τ), while a sequence of consecutive atoms forms a *molecule* (Fig. 2). In this scheme, vibration and silence atoms represent bits '1' and '0', respectively. While naive bit-to-atom mapping is intuitive, commodity vibration

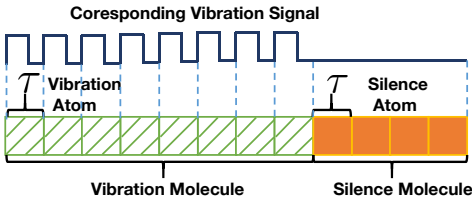

Fig. 2: An example of the symbol in Vib-Sound, which consists of multiple atoms and molecules.

Algorithm 1: Symbol Duration Determination

Input : t^v, t^s
Output: L^{opt}

for L_i *in* L **do**
  for t_j^v *in* t^v **do**
    $p \leftarrow \frac{t_j^v}{\tau}$, $q \leftarrow \frac{L_i - t_j^v}{\tau}$;
    if $t_j^{va} \neq t_j^{vb}$ *for any* t_j^{va}, t_j^{vb} **then**
      $n \leftarrow A_{p+q}^{p}$
    else if $t_j^{va} = t_j^{vb}$ *for any* t_j^{va}, t_j^{vb} **then**
      $n \leftarrow C_{p+q}^{p}$
    else
      $n \leftarrow C_{p+q-c}^{c} A_{p+q-c}^{p-c}$
    $S_{L_i} \leftarrow S_{L_i} + n$
  if $S_{L_i}^{opt} < S_{L_i}$ **then**
    $S_{L_i}^{opt} \leftarrow S_{L_i}$, $L^{opt} \leftarrow L_i$

motors exhibit significant mechanical inertia, leading to non-deterministic hardware latencies. As illustrated in Fig. 3, these manifest as: (i) **modulation delay** (startup latency $t_1 - t_0$) and (ii) **stopping delay** (residual vibration $t_{stop} - t_2$). Such delays, compounded by environmental noise (e.g., floor vibrations), induce significant Inter-Symbol Interference (ISI) and signal distortion. To achieve reliable communication, IVM must strategically aggregate atoms into molecules, optimizing the pulse-width to ensure the modulated symbols remain distinguishable despite these hardware-induced temporal errors.

In Vib-Sound, to improve the throughput, we introduce a ***symbol duration determination*** algorithm to find the optimal symbol duration. We design a ***vibration symbol selection*** scheme to modulate each symbol with desired molecules to minimize intersymbol interference (ISI). The first step is to model the vibration signal. Typically, the sampling rate of the accelerometer is around $1.6kHz$ [15]. However, due to security concerns, recent smart device manufacturers have restricted the sampling rate of accelerometers to $200Hz$ [10], [16]. Therefore, we use the square wave to model the vibration signal. In Vib-Sound, we model the vibration signal as $x(t) = \boldsymbol{D} sign(sin(2\pi f t))$, where f is the motor's vibration frequency and t is the signal duration. $\boldsymbol{D}$ is an integer ($\boldsymbol{D} \in \{0, 1\}$) indicating vibration status. When $\boldsymbol{D}$ is 1, the motor vibrates; otherwise, it is in silence mode. Then, a symbol S composes multiple molecules that can be modeled as: $S(t) = \boldsymbol{D} \odot \text{sign}(\sin(2\pi \boldsymbol{f} t))$, where $\boldsymbol{f} = [f_0, f_1, \ldots, f_n]^T$. Hence, each row represents a molecule, vibration, or silence.

Now, we can use the symbol model to find the optimal symbol duration L. Specifically, we denote the motor's minimal vibration duration as t_{min}^v. Due to the limitations

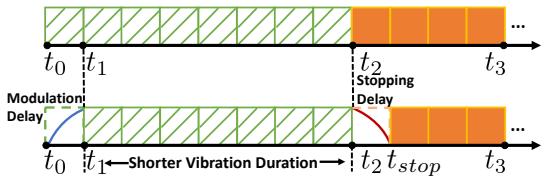

Fig. 3: Errors introduced during the modulation process.

Algorithm 2: Vibration Symbol Selection

Input : L^{opt}, Θ, n
Output: Γ

Generate a random solution Γ^r from Θ, Set R^{opt}
for *iter in* n **do**
  Ruin step: $\Gamma^r \leftarrow \Gamma^r \setminus sub$
  Reconstruct step: $\Gamma^{r2} \leftarrow \Gamma^r \cup sub2$
  $R \leftarrow \min_{S_i, S_j \in \Gamma^{r2}} dis(S_i, S_j) - \min_{S_i, S_j \in \Gamma^r} dis(S_i, S_j)$
  if $R > R^{opt}$ **then**
    $R^{opt} \leftarrow R, \Gamma \leftarrow \Gamma^{r2}$

of the motor, t_{min}^v is much larger than the accelerometer's minimal detectable segment τ. Then, for a symbol S, possible lengths t^v for vibration molecule can be calculated as: $t^v \in \left\{t_{min}^v, ..., t_{min}^v + \tau \lfloor \frac{L - t_{min}^v}{\tau} \rfloor\right\}$. Similarly, the possible lengths t^s for silence molecules can be calculated as: $t^s \in \left\{0, \tau, ..., \tau + \tau \lfloor \frac{L - t_{min}^v - \tau}{\tau} \rfloor\right\}$. Therefore, the symbol duration L should be $L = \sum_{i=0}^{m} t_i^v + \sum_{j=0}^{n} t_j^s$, where m and n are the numbers of vibration and silence molecules, respectively. As we can see from this equation, the longer the symbol duration, the more molecules can be used to modulate symbols. However, the longer the symbol duration, the fewer symbols can be transmitted per second, reducing the throughput. Therefore, we should find optimal symbol length L^{opt} to represent as many symbols as possible while improving the throughput.

To ensure reliable communication, Vib-Sound must determine an optimal symbol duration L^{opt} and a corresponding robust symbol set. Algorithm 1 iterates through combinations of vibration (p) and silence (q) molecules to identify candidates under varying constraints: (i) uniform duration, (ii) unique duration, and (iii) mixed duration molecules. In our implementation, we derive $L^{opt} = 140$ ms as the balance between throughput and reliability. However, selecting a modulation set Γ from the exhaustive space of possible combinations Θ is computationally prohibitive (e.g., $|\Theta| \approx 2^{26}$ for $L^{opt} = 140$ ms). To maximize resilience against hardware jitter and environmental noise, we define the selection criteria as a Hamming distance optimization problem. We aim to find a set $\Gamma \subset \Theta$ that maximizes the minimum distance between any two symbols S_i, S_j:

$$\arg\max_{\Gamma} \min_{S_i, S_j \in \Gamma, i \neq j} D_H(S_i, S_j) \tag{1}$$

where $D_H(\cdot)$ denotes the Hamming distance. This ensures that even with significant bit-level errors caused by motor startup/stopping delays, the receiver can uniquely map a noisy

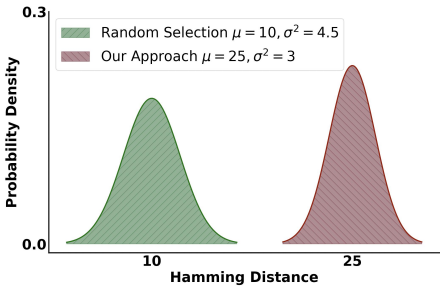

Fig. 4: Hamming distance distributions of Random Selection and our proposed IVM algorithm.

sequence to its nearest valid symbol in Γ.

We introduce a ***vibration symbol selection*** scheme in Algorithm 2 based on a *ruin-and-reconstruct* concept [17], which periodically destroys parts of a solution and rebuilds them to avoid getting stuck in local optima. We randomly select a solution Γ^r from Θ and set the number of iterations to n (lines 1-2). In each iteration, we ruin the initial solution by removing multiple symbols so that the remaining symbols have higher Hamming distances, and reconstructs a new solution Γ^{r2} by randomly inserting new symbols from Θ (lines 3-4). At last, we find the desired solution Γ by comparing the differences between Hamming distances in Γ^r and Γ^{r2} (lines 5-7). We benchmark inter-symbol Hamming distance distribution (normal distribution) of the *Random Selection* $\sim \mathcal{N}(10, 4.5)$ and our proposed *IVM* approach $\sim \mathcal{N}(35, 3)$. Per Figure 4, IVM approach drastically reduces the inter-symbol interference over the Random Selection.

B. Multi-Medium Channel Modeling (MCM)

To mitigate the sensitivity of dual-modality signals, Vib-Sound employs Multi-Medium Channel Modeling (MCM) for dynamic signal compensation. Unlike standard wireless channels, vibration propagation is highly dependent on the physical surface, where the duration of a vibration "molecule" directly alters the surface's mechanical response.

We address this through a payload-dependent preamble (Fig. 5) designed for joint channel estimation and packet synchronization. Each preamble consists of: (i) an **indicator** (two atomic pulses) for detection; and (ii) **boundary molecules**, representing the shortest and longest pulse durations present in the subsequent data packet. By embedding these extreme vibration states, the preamble allows Vib-Sound to dynamically characterize the non-linear channel response and pulse-width distortion inherent in the physical medium, ensuring robust demodulation across diverse surfaces and noise profiles. Let p_v and p_a denote the ideal vibration and acoustic preambles, respectively, constructed using shortest/longest molecules (L_{short}, L_{long}) and atomic pulses (v, a). The received preambles p'_v and p'_a are modeled as:

$$p'_m = H_m p_m + n, \quad m \in \{v, a\} \tag{2}$$

where H_m represents the channel matrix for modality m. Although the microphone's sampling rate (44.1 kHz) far exceeds the accelerometer's (200 Hz), causing a dimension

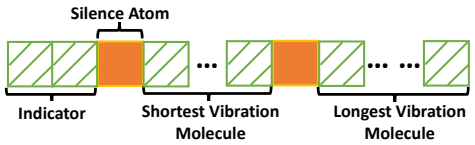

Fig. 5: Preamble contains an indicator, a shortest vibration molecule, and a longest vibration molecule.

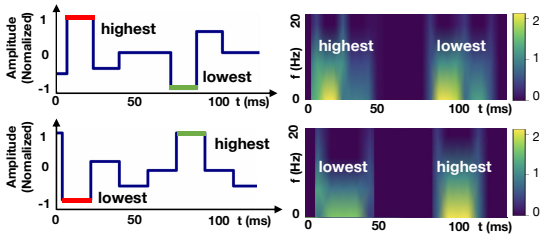

Fig. 6: The receiver will receive different signals even if the sender transmits two identical symbols.

mismatch between p_v and p_a, Vib-Sound estimates each channel independently. Transmitted symbols s_m are then recovered via zero-forcing equalization: $\hat{s}_m = H_m^{-1} s'_m$.

Synchronization Analysis. A potential concern in multi-medium systems is the propagation delay difference between vibration ($V_v \approx 3,968$ m/s [18]) and acoustic signals ($V_a \approx 343$ m/s). However, in typical short-range applications ($d < 0.5$ m), the Time-of-Arrival (ToA) difference is $\Delta t = \frac{d}{V_a} - \frac{d}{V_v} < 1.37$ ms. This gap is negligible as it falls well below the commodity accelerometer's sampling resolution ($\approx$ 5 ms [19]) and is an order of magnitude smaller than our 100 ms preamble. Thus, Vib-Sound achieves robust joint demodulation without explicit inter-modality synchronization. Consequently, the system can treat the dual-modality inputs as quasi-simultaneous, significantly reducing computational overhead by eliminating the need for complex temporal alignment.

C. Collaborative Cross-Channel Demodulation

Collaborative Cross-Channel Demodulation (CCD) aims to achieve real-time demodulation with high reliability while enhancing throughput. Several challenges arise in this process. *First*, due to the hardware limitation and channel noise, the received signal suffers high distortions and uncertainty. For example, as shown in Figure 6, the receiver detects two significantly different signals when the sender transmits two identical symbols. The first signal has high values from 0 to $50ms$ in both time and frequency domains, while the second signal is the opposite of the first signal. *Second*, to improve the demodulation reliability, traditional approaches require each symbol to have a minimal guard interval to avoid intersymbol interference (ISI), which sacrifices the throughput. *Third*, since the receivers in Vib-Sound are smart devices, the proposed demodulation approach should be lightweight. To address these challenges, CCD employs an ensemble model that integrates vibration signals, spectrum information, and side-channel acoustic signals. The overall CCD architecture (Figure 7) consists of three main modules: *(i)* a time-domain

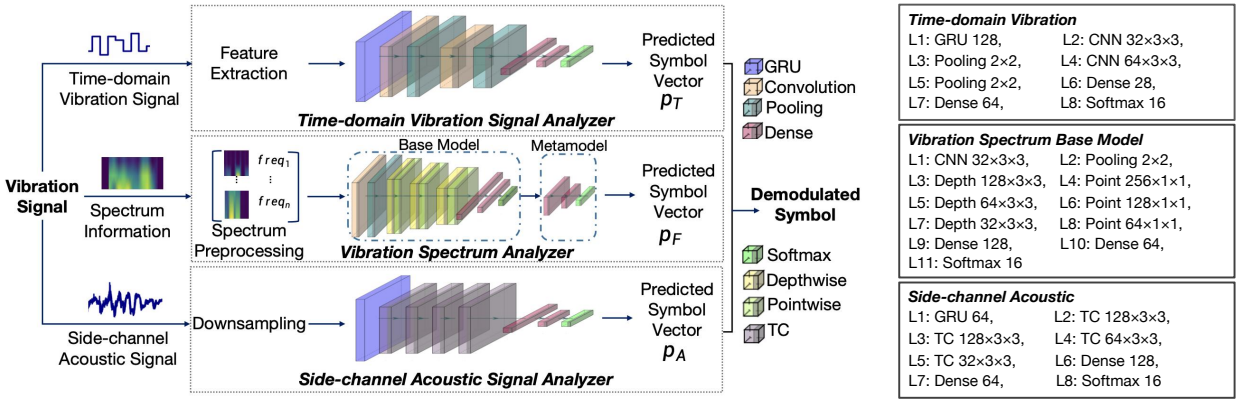

Fig. 7: The architecture of CCD leverages an ensemble model to process the time-domain vibration signal, the spectrum information, and the corresponding acoustic signal simultaneously.

vibration signal analyzer, *(ii)* a vibration spectrum analyzer, and *(iii)* a side-channel acoustic signal analyzer. These modules work together to demodulate each segment of the received vibration signal into a symbol.

The ***time-domain vibration signal analyzer*** processes the received vibration signal in the time domain and outputs the probability distribution of all possible symbols p_T. Due to low sampling rates of commodity accelerometer (*i.e.,* 200Hz), the demodulation results suffer high errors. Hence, we design the time-domain signal analyzer as an eight-layer neural network to improve the data granularity and demodulation accuracy. The signal analyzer includes one GRU layer, two pairs of convolutional layers followed by max-pooling layers, two fully connected ReLU layers, and one fully connected softmax layer, with ReLU activation function. To simplify the analyzer, we use a GRU layer [20] because it has fewer parameters than RNNs [21] and LSTMs [22]. Additionally, due to the low sample rate (200 samples/sec) of vibration signals, we use a convolutional layer for time-domain demodulation.

Given the high attenuation of vibration signals on soft surfaces (e.g., carpet) and ambient noise, demodulating the received signal can be error-prone, especially with data-driven methods. To address this challenge, we propose a novel physics-informed cost function that incorporates elements based on vibration signal principles and accounts for transmission errors, alongside the existing cross-entropy component. (Equation 3).

$$L = -\sum_{i=1}^{M} y_i \log(p_i) + \sum_{j=1}^{3} || e^{-\alpha d_1^j} \int_t^{t+s} Z(t) d_t - \left(\sum_{i=2}^{n} e^{-\alpha d_i^j} \int_t^{t+s} Z(t) d_t + N(t) \right) ||_2 \quad (3)$$

The physics-informed component assesses the discrepancy between the theoretical and practical energy levels of vibration signals and the amount of ambient noise. The loss of the physics-informed component equals the sum loss from three axes. The theoretical energy quantity of vibration signals at axis j is $e^{-\alpha d_1^j} \int_t^{t+s} Z(t) d_t$, where d_1^j is the shortest vibration path from the sender to the receiver at the axis j, s is the vibration duration, and $Z(t)$ is the real vibration signal generated by the sender. The practical energy quantity from other propagation paths alongside the noise quantity equal to $\sum_{i=2}^{n} e^{-\alpha d_i^j} \int_t^{t+s} Z(t) d_t + N(t)$, where $N(t)$ is noisy quantity. The inputs of the time-domain signal analyzer are samples of vibration signals, obtained from the accelerometer in x, y, and z dimensions. Performing feature extraction, we will extract the statistical information (*e.g.,* 30th percentile) and the energy feature (*i.e.,* the offset between the current energy level and the zero energy level) from these samples and feed them into the analyzer to get the predicted probability vector p_T.

In practice, the time-domain vibration signal analyzer sometimes mispredicts demodulated symbols due to limited information (see Figure 6). Given solely relying on time-domain information is insufficient, we propose ***vibration spectrum analyzer*** as CCD's second component for processing the spectrum information of the vibration signal, aiming to enhance the demodulation accuracy. To enable real-time processing on resource-constrained commodity devices, *Vib-Sound* utilizes a lightweight spectrum analyzer architecture inspired by MobileNet's depthwise separable convolutions [23]. The system first transforms raw vibration signals into spectrograms via Short-Time Fourier Transform (STFT), which are partitioned into n sub-frequency bands $[f_1, \ldots, f_n]$. Each band is processed by a dedicated **base model** comprising an initial convolutional layer, max-pooling, and three pairs of convolutional blocks. This specific configuration minimizes computational complexity while maintaining high feature extraction accuracy. To robustly integrate these isolated frequency-domain insights, we employ a meta-learning ensemble approach [24]. Predictions from the base models are fed into a three-layer meta-model—consisting of two L2-normalized fully connected (FC) layers and a softmax output–to generate a refined probability distribution p_F for final symbol estimation.

To mitigate vibration-only demodulation errors, *Vib-Sound* leverages the motor's acoustic side-channel. By utilizing the 44.1 kHz sampling rate of commodity microphones, the system extracts high-resolution temporal features that complement low-bandwidth vibration data. Inspired by Temporal Convolutional Networks (TCN) [25], our analyzer (Fig. 7) captures long-range dependencies using a hybrid architecture:

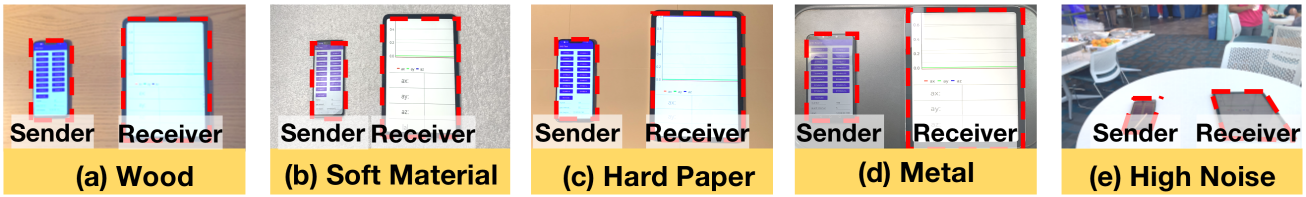

Fig. 8: The implementation of Vib-Sound: (a)–(d) Evaluation across diverse surfaces (wood, cotton, hard paper, metal); (e) Performance under varying noise levels.

one GRU layer, four dilated temporal-convolutional (TC) layers, and two FC ReLU/softmax layers. To ensure feasibility on resource-constrained devices, we employ 1D dilated convolutions, which are more efficient than standard CNNs for high-dimensional features. Furthermore, we implement an **adaptive downsampling** scheme that dynamically adjusts the sampling rate based on instantaneous signal amplitude and frequency, significantly reducing computational overhead before generating the probability distribution p_A. To address acoustic distortion and multi-path attenuation, we introduce a physics-informed cost function (Eq. 4) that augments standard cross-entropy loss with a regularization component tailored for acoustic propagation:

$$L = -\sum_{c=1}^{M} y_c \log(\hat{p}_c) + |A_1 e^{-j\theta_1} Z(t) - \sum_{i=2}^{L} A_i e^{-j\theta_i} Z(t)| \quad (4)$$

This component minimizes the disparity between the theoretical shortest-path energy ($A_1 e^{-j\theta_1} Z(t)$) and the aggregated energy of all alternative $L-1$ acoustic paths ($\sum_{i=2}^{L} A_i e^{-j\theta_i} Z(t)$), where A_i and θ_i represent the attenuation and phase factors, respectively.

Final Fusion. *Vib-Sound* employs soft voting across the probability distributions p_T, p_F, and p_A to identify the most probable symbol. It is optimized by our custom loss function, which jointly maximizes segment-level energy prediction and global classification accuracy, ensuring robust performance across varying environmental conditions (Figs. 9–12).

V. Implementation and Evaluation

A. Implementation

To fully study the robustness of our design, we implement the sender part of Vib-Sound on Huawei Nova 3. The motor in this smartphone only supports on-and-off control, and it can only generate vibration signals with extremely low energy (lower than 4 dB). We use an iPad (10th generation) as the receiver for demodulation. The sampling rate of the accelerometer in an iPad is 200 Hz, and the minimal detectable length of the vibration signal is 5 ms. Therefore, according to the Interference-Aware Vibration Modulation in our design, the optimal symbol length is set to 140 ms. We extensively evaluate Vib-Sound under various settings and scenarios, including different surfaces, including Wood, Soft Material (Cotton), Hard Paper (Cardboard), and Metal with various noise levels, as shown in Figure 8. We vary the communication distances between the sender and receiver from 0 cm to 40 cm and show the experiment results averaged over 100 runs. We also want to mention that during the evaluation, the average demodulation time for each symbol is only 3.62 ms, and the memory usage for the receiver (iPad) is less than 10.5 MB.

B. Evaluation Results

*1) **Throughput vs. Distance**:* We first study the throughput of Vib-Sound under different communication distances in Figure 9 in the lab scenario. The overall throughput of Vib-Sound on metal is higher than that of the remaining surfaces. In this scenario, Vib-Sound achieves the throughput as high as 70 bps, almost ***doubling*** the throughput compared to the most recent vibration-based communication solution (40 bps). On the contrary, the throughput of Vib-Sound is reduced when the sender and receiver are on the soft material, primarily since soft material absorbs vibration signal, hindering the detection of the transmitted symbols. However, we argue that the motor in the sender can only generate extremely low vibration signals, allowing Vib-Sound to achieve 59 bps when the communication distance is 10 cm. **In summary,** Vib-Sound can achieve high throughput and show relatively high robustness on different surfaces.

*2) **Symbol Reception Ratio vs. Distance**:* Figure 10 shows the communication reliability of Vib-Sound under different distances in the lab scenario. When the communication distance is close to 0 cm, the symbol reception ratios are as high as 98% (Metal), 91% (Wood), 89% (Hard Paper), and 85% (Soft Material). As the communication distances increase, the symbol reception ratios reduce. We also need to mention that the energy of the vibration signal generated by the sender is less than 4 dB. However, when the communication distance is as high as 30 cm, the symbol reception ratios on Wood and Metal are still higher than 70%, proving the effectiveness of our design. **In summary,** Vib-Sound can achieve high communication reliability on different surfaces.

*3) **System Sensitivity Analysis**:* This subsection studies the performance of Vib-Sound under high-noise scenarios and different system settings.

Different Noise Levels. Figures 11 and 12 show the performance of Vib-Sound under different noise levels. In these experiments, we set the communication distance to 10 cm, and vary the noise levels (*i.e.,* 33 dB, 55 dB, 65 dB, and 70 dB) according to NPC and CDC noise standards to mimic different scenarios (*i.e.,* quiet room, bathroom with exhaust fan, office room with a printer on, and room with clothes washer on). The results show that Vib-Sound has a relatively stable performance under high noise levels (*e.g.,*

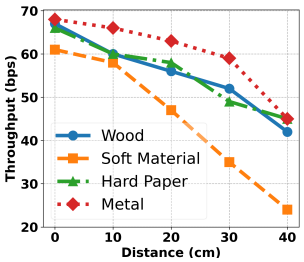

Fig. 9: Throughput vs. Dist.

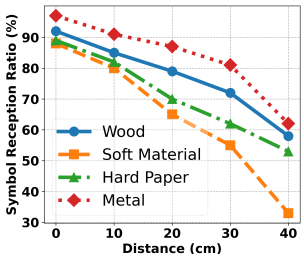

Fig. 10: Reception Ratio vs. Dist.

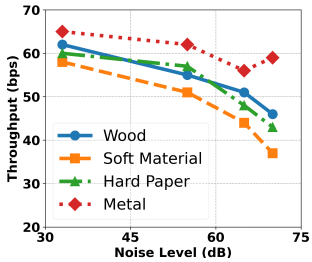

Fig. 11: Throughput vs. Noise

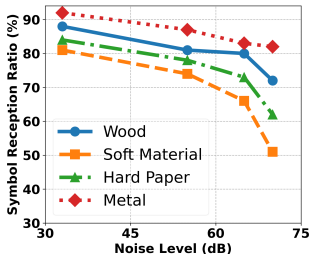

Fig. 12: Reliability vs. Noise

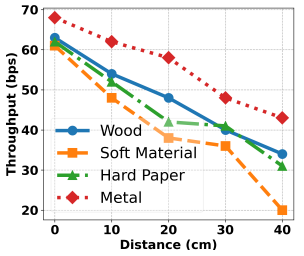

Fig. 13: Throughput vs. Distances (Sender's Screen facing towards the surface)

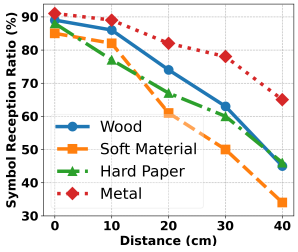

Fig. 14: Reliability vs. Distances (Sender's Screen facing towards the surface)

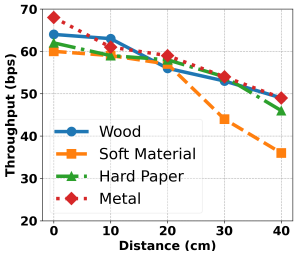

Fig. 15: Throughput vs. Distances (Receiver's Screen facing towards the surface)

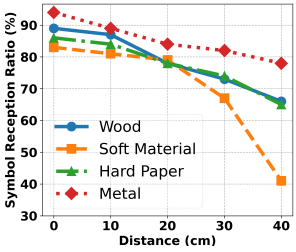

Fig. 16: Reliability vs. Distances (Receiver's Screen facing towards the surface)

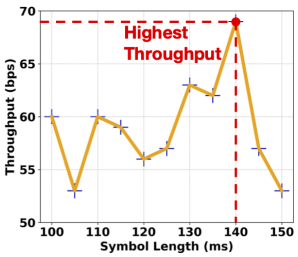

Fig. 17: Impact of Symbol Duration Algorithm.

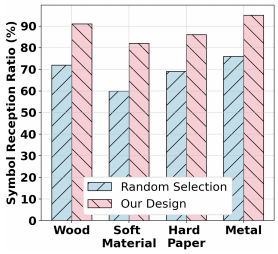

Fig. 18: Impact of Vibration Symbol Selection Algorithm

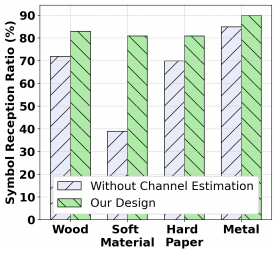

Fig. 19: Impact of Multi-Medium Channel Modeling

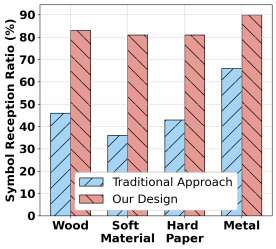

Fig. 20: Demodulation Approaches Comparison

65 dB). For instance, Vib-Sound's performance on metal and wood surfaces remains high enough to support fast and reliable communication even as noise levels reach 70 dB. Conversely, Vib-Sound's performance on hard paper and soft materials remains stable when noise levels are below 65 dB. As the noise level increases to 70 dB, it becomes hard for the receiver to detect useful vibration signals. However, we want to mention that even in this scenario, the throughput of Vib-Sound is still around 40 bps, which is the same as the best performance (40 bps) of the most recent vibration communication solution [26]. **In summary,** the performance of Vib-Sound is relatively stable under different noise levels.

Different Orientations. Figures 13 and 14 study the performance of Vib-Sound when the sender's screen is facing toward the surface, with a set distance of 10 cm. In the experiments, since the flat screen of the smartphone will touch the surface, it is hard for the motor to make the surface vibrate. As a result, the vibration signal suffers high distortions. However, counter-intuitively, compared with Figures 9 and 10 (sender's screen facing up), the throughput and reliability of Vib-Sound are almost the same. This is because Vib-Sound not only leverages the vibration signal but also utilizes the side-channel acoustic signal to conduct demodulation. As a result, Vib-Sound's performance is relatively stable. Figures 15 and 16 show the throughput and symbol reception ratio of Vib-Sound when the receiver's screen faces toward the surface. In this scenario, the vibration signal has to pass through the thick touch panel and the inner screen to reach the accelerometer on the logic board, making it harder for the accelerometer to detect the vibration signal. However, similar to Figures 13 and 14, since the symbols in Vib-Sound are robust against noises, and the demodulation part uses an ensemble model, the Vib-Sound's performance is not affected by the orientation of the receiver. **In summary,** Vib-Sound is robust regardless of the sender or receiver orientation.

*4) **System Insight Analysis**:* This subsection will answer why Vib-Sound can achieve fast and reliable vibration communication. We set the distance to 10 cm, the noise level to 33 dB, and devices facing upward. Since all experiments show a similar trend, we only show the results of the metal surface.

TABLE I: Participants Biometric Profiles

Gender	Age Range	Height	Weight
Male: 12 Female: 8	<10: 3 10-20: 3 20-30: 12 30-40: 2	0-100cm: 3 100cm-160cm: 3 160cm-170cm: 4 170cm-180cm: 7 180cm-200cm: 3	<100lb: 3 100lb-150lb: 4 150lb-180lb: 10 >180lb: 3

Symbol Duration Determination Algorithm. In previous experiments, we set the optimal symbol length to 140 ms. As depicted in Figure 17, we will study the effectiveness of our Symbol Duration Determination algorithm. When the symbol length increases from 100 ms to 140 ms, the throughput increases nonlinearly. When the symbol length reaches 140 ms, Vib-Sound's throughput reaches 68 bps, but drops rapidly as the symbol lengths increase to 150 ms. This experiment proves the effectiveness of our symbol duration determination algorithm. **In summary,** the Symbol Duration Determination algorithm can find the optimal symbol duration for high throughput communication.

Vibration Symbol Selection Algorithm. Figure 18 shows the effectiveness of our Vibration Symbol Selection algorithm. In this experiment, we implement a vibration communication system that randomly selects different combinations of molecules for symbol modulation. Evidently, our design shows dominant advantages compared to the random selection approach. The symbol reception ratios of Vib-Sound are around ***2*** times higher than that of the random selection in all the cases (*i.e.,* wood, soft material, hard paper, and metal). **In summary,** the Vibration Symbol Selection algorithm can find the desired modulation schemes and significantly improve the vibration communication reliability.

Multi-Medium Channel Modeling. In Figure 19, we evaluate the performance of our Multi-Medium Channel Modeling scheme. We can observe that modeling the vibration communication and the acoustic communication channels improved Vib-Sound's performance. For the metal surface, since the motor can easily make the metal vibrate, the vibration signals transmitted from the sender have relatively low errors. In this case, our Multi-Medium Channel Modeling can achieve around 5.8% symbol reception ratio improvement. However, the vibration and side-channel acoustic signals are severely absorbed and distorted for the soft material surface, introducing high errors to the transmitted signals and resulting in a low symbol reception ratio of 39% without channel modeling. In contrast, our channel modeling approach achieves a symbol reception ratio of 81%, effectively ***doubling*** the symbol reception ratio. **In summary,** the Multi-Medium Channel Modeling scheme can effectively model the communication channels and conduct signal compensation on different surfaces.

Collaborative Cross-Channel Demodulation. Figure 20 show the effectiveness of our demodulation approach. In this experiment, we use the maximum likelihood demodulation as the state-of-the-art solution. Evidently, our Collaborative Cross-Channel Demodulation approach significantly improves the symbol reception ratio. While our approach boosts the symbol reception ratio on metal surfaces from 67% to a remarkable 90%, it achieves an even more impressive improvement on soft material surfaces, raising the symbol reception ratio from 37% to an outstanding 81.5%. This is because our demodulation approach takes the time-domain vibration signal, the corresponding spectrum information, and the side-channel acoustic signal as the input, which increases the data granularity. **In summary,** our Collaborative Cross-Channel Demodulation approach can significantly improve the vibration communication reliability.

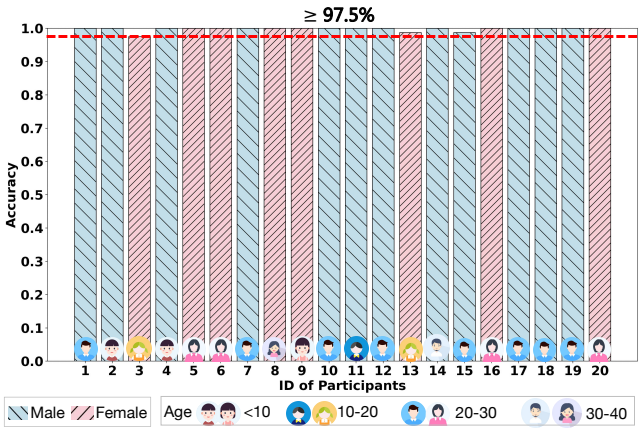

Fig. 21: Vib-Sound's authentication performance. The results suggest an average accuracy of 99.75%.

C. Authentication Use Case

To demonstrate the practicality of Vib-Sound's design, we developed a proof-of-concept multi-factor authentication application. The second factor involves the user's smartphone transmitting a one-time code to the receiver using Vib-Sound's communication system, while the third factor incorporates human biometrics. Specifically, during the transmission of the one-time code via the vibration medium, the user places their finger on the smartphone, affecting the vibration signal with their distinct body mass. We conducted an experiment with 20 participants of various anthropometric profiles (Table I) to validate the effectiveness of our authentication system across different demographics. By including individuals of different ages, genders, heights, and weights, we ensure that our system is robust and inclusive, demonstrating Vib-Sound's practical applicability and reliability.

Considering the authentication application runs on individual participants' devices, we trained a unique binary classifier for each participant to identify whether the subject is the authorized participant. We then evaluate the accuracy of individual applications by providing a mix of samples, including both authorized and unauthorized attempts. As depicted in Figure 21, seventeen out of the twenty authentication applications achieved perfect accuracy of 100%, with the lowest accuracy recorded at 97.5%, resulting in an overall accuracy of 99.75%. These results demonstrate the efficacy of Vib-Sound's design in achieving robust authentication in real-world scenarios.

VI. Related Work

Using vibration and acoustic signals to support various applications, such as communication [1], [2], authentication

[7], [27], and identification [10], [28], has become one of the most exciting topics recently. For example, to support communication, pioneer work – Ripple [1] first investigates the possibility of vibration-based communication by detecting the vibration signals using a customized accelerometer. Built on top of Ripple [1], Ripple II [2] further improves the communication speed by using a customized microphone receiver. However, these approaches require customized devices, which may limit the potential applications. To overcome this challenge, LeaD [26] applies machine learning to support vibration communication with a bit rate of $40bits/s$ using commodity devices. Recently, several works have shown that vibration signals can support two-factor authentication for smart devices [7], [27]. Moreover, researchers have even proven that the vibration signal can be used to measure and identify liquids' viscosity [10]. Unlike existing approaches, Vib-Sound is the first work that achieves fast and reliable communication using vibration and side-channel acoustic signals simultaneously. Our design does not require any hardware modifications and can be directly implemented on commodity devices, which can be widely applied to support real-world applications.

VII. Conclusion

This paper presents Vib-Sound, the first fast and reliable vibration communication system using commodity devices, such as smartphones and tablets. To do this, we introduce multiple novel techniques, including the *(i)* Interference-Aware Vibration Modulation (IVM) scheme to determine the optimal symbol length and generate the desired vibration signal, the *(ii)* Multi-Medium Channel Modeling (MCM) scheme to perform channel modeling and signal compensation, and the *(iii)* Collaborative Cross-Channel Demodulation (CCD) approach. The CCD approach employs a novel physics-informed learning-based technique to simultaneously demodulate the vibration and corresponding side-channel acoustic signals generated by the motor. Our extensive evaluation indicates that Vib-Sound can achieve fast and reliable communication under various settings and scenarios. We also believe that Vib-Sound opens a new direction for enabling the collaboration between vibration and acoustic signals using commodity devices.

References

[1] N. Roy, M. Gowda, and R. R. Choudhury, "Ripple: Communicating through physical vibration," *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2015.

[2] N. Roy and R. R. Choudhury, "Ripple ii: Faster communication through physical vibration," *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2016.

[3] K. Cui, Q. Yang, Y. Zheng, and J. Han, "Mmripple: Communicating with mmwave radars through smartphone vibration," *ACM IPSN*, 2023.

[4] J. Liu and B. Chen, "Sonicsense: Object perception from in-hand acoustic vibration," in *Conference on Robot Learning*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270738210

[5] F. Li, J. Zhao, H. Yang, D. Yu, Y. Zhou, and Y. Shen, "Vibhead: An authentication scheme for smart headsets through vibration," *ACM Transactions on Sensor Networks*, vol. 20, pp. 1 – 21, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259286872

[6] R. Wijewickrama, S. Dohadwalla, A. Maiti, M. Jadliwala, and S. Narain, "Skinsense: Efficient vibration-based communications over human body using motion sensors," *Internet Things*, vol. 23, p. 100835, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257482438

[7] E. Husa and R. Tourani, "Vibe: An implicit two-factor authentication using vibration signals," *IEEE CNS*, 2021.

[8] S. Lee, W. Choi, and D. H. Lee, "Usable user authentication on a smartwatch using vibration," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 304–319.

[9] J.-W. Chang, K. Sun, D. Xia, X. Zhang, and F. Koushanfar, "Eveguard: Defeating vibration-based side-channel eavesdropping with audio adversarial perturbations," *2025 IEEE Symposium on Security and Privacy (SP)*, pp. 4534–4552, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:274116881

[10] Y. Huang, K. Chen, Y. Huang, L. Wang, and K. Wu, "Vi-liquid: unknown liquid identification with your smartphone vibration," in *ACM MobiCom*, 2021.

[11] A. Lorenz, B. Siewertsen, V. K. Clemmensen, J. B. Petersen, J. Friederich, and S. Lazarova-Molnar, "Vibration data analysis for fault detection in manufacturing systems-a systematic literature review," in *IEEE ICIEA*, 2022.

[12] R. Xiao, S. Mayer, and C. Harrison, "Vibrocomm: Using commodity gyroscopes for vibroacoustic data reception," in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2020.

[13] W. Wang, T. Xie, X. Liu, and T. Zhu, "Ect: Exploiting cross-technology concurrent transmission for reducing packet delivery delay in iot networks," in *IEEE INFOCOM*, 2018.

[14] W. Wang, X. Liu, Y. Yao, Y. Pan, Z. Chi, and T. Zhu, "Crf: Coexistent routing and flooding using wifi packets in heterogeneous iot networks," in *IEEE INFOCOM*, 2019.

[15] G. Grouios, E. Ziagkas, A. Loukovitis, K. Chatzinikolaou, and E. Koidou, "Accelerometers in our pocket: Does smartphone accelerometer technology provide accurate data?" *Sensors*, 2022.

[16] W. Song, H. Jia, M. Wang, Y. Wu, W. Xue, C. T. Chou, J. Hu, and W. Hu, "Pistis: Replay attack and liveness detection for gait-based user authentication system on wearable devices using vibration," *IEEE Internet of Things Journal*, 2022.

[17] T.-S. Khoo, B. B. Mohammad, V.-H. Wong, Y.-H. Tay, and M. Nair, "A two-phase distributed ruin-and-recreate genetic algorithm for solving the vehicle routing problem with time windows," *IEEE Access*, 2020.

[18] M. Krause, U. Dackermann, and J. Li, "Elastic wave modes for the assessment of structural timber: Ultrasonic echo for building elements and guided waves for pole and pile structures," *Journal of Civil Structural Health Monitoring*, 2015.

[19] B. Groza, A. Berdich, C. Jichici, and R. Mayrhofer, "Secure accelerometer-based pairing of mobile devices in multi-modal transport," *IEEE Access*, 2020.

[20] K. C. Junyoung Chung, Caglar Gulcehre and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS*, 2014.

[21] R. M. Schmidt, "Recurrent neural networks (rnns): A gentle introduction and overview," *ArXiv*, vol. abs/1912.05911, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:209324034

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997. [Online]. Available: https://api.semanticscholar.org/CorpusID:1915014

[23] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CVPR*, 2017.

[24] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *PMLR*, 2017.

[25] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[26] G. Zhao, B. Du, and Shen, "Lead: Learn to decode vibration-based communication for intelligent internet of things," *ACM TOSN*, 2021.

[27] X. Xu, J. Yu, Y. chen, Q. Hua, Y. Zhu, Y.-C. Chen, and M. Li, "Touchpass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations," *ACM MobiCom*, 2020.

[28] S. A. Anand, J. Liu, C. Wang, M. Shirvanian, N. Saxena, and Y. Chen, "Echovib: Exploring voice authentication via unique non-linear vibrations of short replayed speech," in *Proceedings of the ACM ASIACCS*, 2021.