Deaf-Aid: Mobile IoT Communication Exploiting Stealthy Speaker-to-Gyroscope Channel

Ming Gao
Zhejiang University
gaomingppm@gmail.com

Feng Lin*
Zhejiang University
flin@zju.edu.cn

Weiye Xu
Muertikepu Nuermaimaiti
xuweiye@zju.edu.cn
murtakip@zju.edu.cn
Zhejiang University

Jinsong Han
Zhejiang University
hanjinsong@zju.edu.cn

Wenyao Xu
SUNY Buffalo
wenyaoxu@buffalo.edu

Kui Ren
Zhejiang University
kuiren@zju.edu.cn

ABSTRACT

Internet of Things (IoT) devices are hindered from communicating with their neighbors by incompatible protocols or electromagnetic interference. Existing solutions adopting physical covert channels have limitations in receiver distinction, additional hardware, conditional placement, or physical contact. Our system, *Deaf-Aid*, utilizes the stealthy speaker-to-gyroscope channel to build robust protocol-independent communication with automatic receiver identification. *Deaf-Aid* exploits ultrasonic signals at a frequency corresponding to the target receiver, forcing the gyroscope inside to resonate, so as to convey information. We probe the relationship among axes in a gyroscope to surmount frequency offset ingeniously and support multi-channel communication. Meanwhile, *Deaf-Aid* identifies the receivers automatically via device fingerprints constituted by the diversity of resonant frequency ranges. Furthermore, we entitle *Deaf-Aid* the capability of mobile communication which is an essential demand for IoT devices. We address the challenge of accurate signals recovery from motion interference. Extensive evaluations demonstrate that *Deaf-Aid* yields 47bps with BER lower than 1% under motion interference. To our best knowledge, *Deaf-Aid* is the first work to enable stealthy mobile IoT communication on the basis of inertial motion sensors.

CCS CONCEPTS

• **Hardware → Sensor applications and deployments**.

KEYWORDS

gyroscopes, covert channel, mobile IoT communication

*Feng Lin is the corresponding author.

MobiCom '20, September 21–25, 2020, London, United Kingdom

ACM ISBN 978-1-4503-7085-1/20/09…$15.00
https://doi.org/10.1145/3372224.3419210

ACM Reference Format:
Ming Gao, Feng Lin, Weiye Xu, Muertikepu Nuermaimaiti, Jinsong Han, Wenyao Xu, and Kui Ren. 2020. Deaf-Aid: Mobile IoT Communication Exploiting Stealthy Speaker-to-Gyroscope Channel. In *The 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20), September 21–25, 2020, London, United Kingdom.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3372224.3419210

1 INTRODUCTION

Internet of Things (IoT) has attracted increasing attention in recent years. It connects various electronic appliances intelligently for people's convenience. Analysts predict that the expenditure on the deployment of IoT will continue to maintain good momentum, rising to $726.5 billion worldwide annually [24]. Artificial intelligence and 5G communication technology also help to combine various devices, aiming at building a comprehensive IoT network.

However, creating such an everything-related IoT network involves abundant obstacles. Various incompatible communication standards have aggravated the problem during information exchange via IoT devices. Requirements in different scenarios encourage various protocols, while devices usually support only one or a couple of them. Wi-Fi [27] and Bluetooth [62] are widely used in mobile communication. ZigBee [52] and MQTT [22] are suitable for the transmission of small-streamed data, especially under resource constraints. Furthermore, there are EnOcean [38], 6LowPan [59] in the field of smart home and AMQP [49], COAP [7] in industrial IoT. To make matters worse, manufacturers develop their own protocols, building distinctive systems to attract consumers. These aforementioned methods rely only on the electromagnetic wave and would fail upon the electromagnetic interference and shielding, as demonstrated in Fig. 1.

To address these problems above, researchers take advantage of physical characteristics to build a covert channel between nodes that are physically and logically separated [28, 56, 57], so that devices can communicate regardless of the protocols. Nevertheless, these systems are confronted with several hindrances, such as additional hardware, confined placement, or physical contact. For instance, *Ripple* [40, 41] demands specialized vibration motors and physical contact; *BitWhisper* [17] can only be applied between two desktop PCs in fixed position. Moreover, they are dependent on

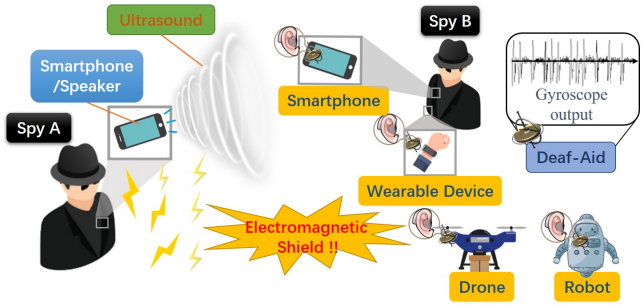


Figure 1: An application scenario for *Deaf-Aid* in the case of espionage. Spy A divulges privacy secretly to Spy B, who receives it stealthily via various IoT devices under the electromagnetic shield.

manual receiver identification, which is impractical in a comprehensive and mobilizable IoT network. More feasible and robust communication between IoT devices is needed urgently [48].

We turn attention to the channel of speaker-to-gyroscope. It has been interpreted that micro-electro-mechanical systems (MEMS) inertial sensors are vulnerable to the ultrasonic injection [42, 47, 48, 51]. Choreographed ultrasound can couple to the stationary MEMS gyroscopes and make them produce low-frequency angular rate readings [10, 11]. However, little attention was drawn on the potential benefits of its susceptibility. Inspired by it, we explore gyroscopes resonance from a communication perspective. Despite the limitation of protocols, a robust system is proposed for bridging a stable transmission in an IoT network, transmitting via speakers, and decoding them through gyroscopes. The channel frequency is selected according to the receivers as each gyroscope has its own unique resonant frequency range. Such a non-contact speaker-to-gyroscope channel in IoT communication is feasible because ultrasonic signals can be obtained through commodity speakers in phones or voice assistants without any peripherals, and gyroscopes have become an indispensable part in intelligent devices [53], including smartphones, VR sets, vehicles, wearable devices, remote control devices and the like.

Robust communication among mobilizable devices is significant. Movement introduces noise, masking characteristic signals, especially on the gyroscope-based system. Robustness to motion is certainly a key issue in inertial sensors reutilization [9, 29, 35, 50, 58]. Moreover, unpredictable frequency offset confuses the frequency characteristics, hindering accurate signals recovery via spectrum analysis. In this case, the gyroscope-based systems are unable to work stably and precisely in a dynamic environment.

For robust gyroscope-based communication in these circumstances, our system needs to specifically address several practical challenges: (1) **Capability**: How to leverage gyroscopes to build a protocol-independent channel of high quality with precise receiver identification. (2) **Mobility**: How to accurately recover the signals in a mobile communication scene. (3) **Drift**: How to deal with the frequency offset caused by drift to ensure communication stability.

To this end, we present a convenient and robust system that exchanges data over the air, namely *Deaf-Aid*, free from restrictions including peripherals, fixed positions, and artificial receiver identification. It provides an alternative and complementary communication channel to current IoT devices. We model the resonant output of gyroscopes, analyze the frequency offset, and exploit the inter-axial relation for correction. The compositions of *Deaf-Aid* are elaborated, including receiver identification, encoding, denoising, and threshold. It supports simultaneous communication on double channels, even from two transmitters. Movement influence is taken into account and described exhaustively. Multiple technologies are employed to adjust our system to a mobile IoT network. We evaluate our system on three kinds of speakers, 32 gyroscopes chips of four models, and three kinds of phones, and exert a variety of movement on it for robustness test. To better evaluate our system, we perform a comprehensive evaluation with 22 participants to validate the effectiveness under real-world scenarios.

The contribution of *Deaf-Aid* can be summarized as follows:

- We investigate the possibility of communicating through a gyroscope and elaborate a stealthy channel without the restriction of peripheral, contact, fixed placement, and especially the manual receiver identification.
- We comprehensively analyze the relationship among axes in a gyroscope under resonance, which has not been studied before in existing literature. Accordingly, noise is removed and motion interference is suppressed even when drift brings about frequency offset.
- We develop a robust communicating system for a mobilizable IoT network. In particular, we take the initiative in excavating the potential of inertial sensors reutilized for the robustness to movement.

2 BACKGROUND

2.1 Gyroscope Structure

MEMS gyroscopes are implemented with Coriolis force [8]. It is usually equipped with movable parts in two orthogonal directions to generate Coriolis force. As shown in Fig. 2, in the driving direction, driving springs add a sinusoidal voltage to force the mass to oscillate at its natural frequency. The Coriolis sensing fingers move owing to the transverse Coriolis motion. In sensing direction, Coriolis acceleration leads to capacitance change. This acceleration a_x is similar to angular rate ω, according to

$$a_x = -2\omega\dot{y}, \tag{1}$$

where $\dot{y}$ is the linear velocity in driving direction. It converts the angular rate into the displacement in sensing direction.

Meanwhile, only one module cannot distinguish between translation and rotation. Consequently, two identical structures are placed

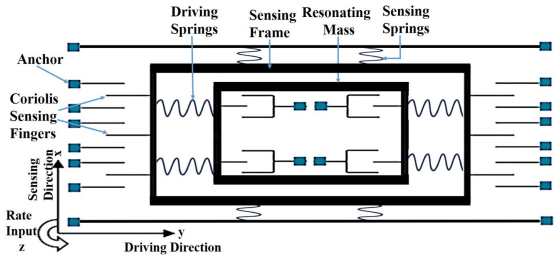


Figure 2: Concept of MEMS gyroscope structure.

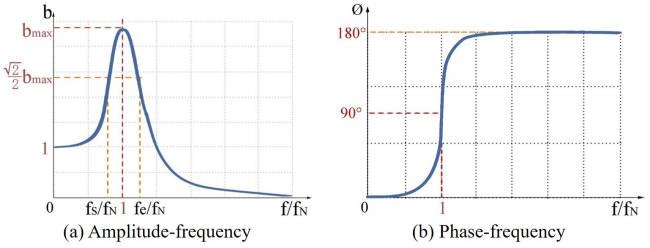


Figure 3: The characteristics of the forced vibration.

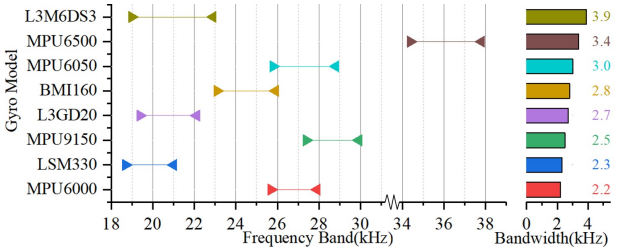


Figure 4: Popular gyroscopes have a narrow resonant frequency range above human audibility.

abreast, linked to a differential amplifier, and then the gyroscope reading is finally obtained after the processes of amplifier, filter, and analog-to-digital conversion.

2.2 Resonance Principle

The structure of MEMS gyroscopes is a kind of single-degree-of-freedom system with a high damping ratio ξ typically. Damping is ignored at low frequency, and gyroscopes keep linear outputs. As frequency increases, damping gradually becomes dominant, and oscillation occurs, with the characteristics of the forced vibration [46] illustrated in Fig. 3. They indicate that the gain coefficient b reaches the peak at the natural frequency f_N, where phase Φ changes dramatically. As a result, gyroscopes respond to acoustic injection. For the sake of accurate measurement, this architecture is designed to share the same natural frequency with resonating mass. However, inevitable errors bring about natural frequency alternation in batch production. This implies the inter-individual discrepancy in the natural frequency among gyroscopes.

3 FEASIBILITY INVESTIGATION

On the basis of characteristics of the ultrasonic wave, we demonstrate and evaluate the feasibility and stealth of the speaker-to-gyroscope channel in respect of noiselessness, availability, and inaudibility. We tested eight models of chips for four each, whose resonant frequency bands are displayed in Fig. 4.

Inaudibility. It has been proved that the resonant frequency of gyroscopes tends to exceed 18kHz in Fig. 4. It is scarcely perceptible to human hearing [10, 11, 42, 51] and always ignored by the speech recognition system, whose sampling rate is below 16kHz. Although an accelerometer also resonate with acoustic injection [6], it is discarded for its audible resonant frequency, within 10kHz usually.

No peripheral. According to the Nyquist sampling theorem, commercial speakers can induce sound within 24kHz with the 48kHz sampling rate. Hi-Fi speakers [54] and contemporary mobiles [14] perform better. For example, Samsung Galaxy S8 is manufactured with a sound card up to 32-bit/384kHz [15]. A smartphone or a commodity speaker can cover the frequency band of most popular gyroscopes and apply *Deaf-Aid* without any peripherals.

Little environment interference. Common application scenarios of ultrasound prefer frequency bands above 40kHz, such as cleaning, medical examination, and treatment. The resonant frequencies of gyroscopes are almost in the band between 18kHz and 40kHz, where few devices work. Therefore, *Deaf-Aid* is shielded from environmental noise. Conversely, the transmission will not affect the normal operations of its surrounding devices.

4 MODEL

In this section, we utilize an axis as an example to develop a physics-based mathematical model to quantitatively analyze the resonant outputs of a gyroscope.

4.1 Oscillation

Acoustic inputs of specific frequency bring harmonic excitation on a gyroscope. Sound waves produce forces of the same frequency on a single-degree-of-freedom system. Hence, the force imposed on one axis can be described by

$$F(t) = A \cdot sin(2\pi f_0 t + \phi_0), \tag{2}$$

where A is the magnitude decided by intensity and position of the sound source, f_0 is the frequency of the sound source, and ϕ_0 indicates the initial phase. In a single-degree-of-freedom system [46], the resulting oscillation is

$$R_0(t) = bA \cdot sin(2\pi f_0 t + \phi_0 + \phi_1), \tag{3}$$

where the gain coefficient b and phase ϕ_1, introduced by resonance, are determined by the frequency ratio f_0/f_N.

4.2 Digitization

Typical MEMS architecture in a gyroscope comprises three parts: amplifier, filter, and analog-to-digital conversion.

Amplifier and Filter: Two identical structures are designed in typical MEMS gyroscopes to obtain rotation via a differential amplifier, and then a low-pass filter (LPF) is aimed at removing noise. An ideal LPF can completely remove the high-frequency noise beyond the cut-off frequency. However, filters are less effective at handling noise whose frequency is much higher than the cut-off frequency in gyroscopes. Even so, it may introduce amplitude alteration besides slight frequency changes and phase shifts. In general, the analog signal in gyroscopes follows this formula,

$$R(t) = bLA \cdot sin(2\pi f_0' t + \Phi), \tag{4}$$

where L is the influence of filter and depends on the output-data-rate (ODR) and insensitive to the variations of f_0 experimentally; f_0' is the frequency under the influence of filter and $\Phi = \phi_0 + \phi_1 + \phi'$ is the overall phase shift while ϕ' is introduced by amplifier and filter. Since L and ϕ' rely on the structural parameters in gyroscopes, they can be regarded as constants in a given gyroscope.

Analog-to-digital Conversion: The frequency of acoustic input is usually over 18kHz, much higher than the sampling rate in

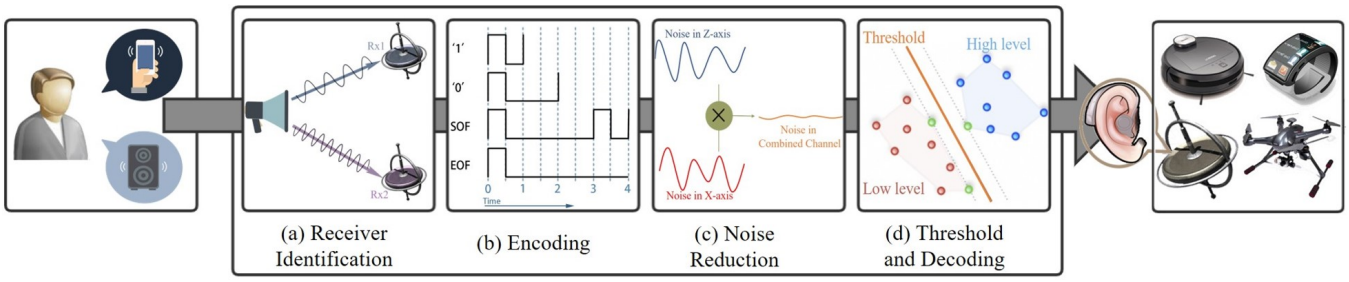


Figure 5: *Deaf-Aid*, a speaker-to-gyroscope channel for mobile IoT communication, where the transmitter is realized by a smartphone or a commercial speaker and the receiver can be any IoT device equipped with a gyroscope.

hundreds. This leads to aliasing, where the sampled signal fails to maintain the original spectrum characteristics. Assuming the sampling rate is Fs, the sampled signal can be expressed as

$$f_0' = n \times Fs + f_1, \quad (-\frac{Fs}{2} < f_1 < \frac{Fs}{2}) \tag{5}$$

$$R[k] = bLA \cdot sin(2\pi f_1 \frac{k}{Fs} + \Phi). \tag{6}$$

In conclusion, we expound the formation of the angular rate readings of gyroscopes under acoustic injection when the gyroscope is stationary. Vividly, the final readings are dependent on the input frequency and sample rate, where aliasing brings about low-frequency readings.

5 FREQUENCY OFFSET CORRECTION

Sample rate drift occurs casually and generates unpredictable frequency offset, making it difficult to separate signals from noise and motion inference via spectrum analysis. In this section, we elaborately analyze this process and exploit the relationship among axes for offset correction.

5.1 Sample Rate Drift

A serious weakness of sampling rate drift is that it leads to obvious but unpredictable deviations of output frequency [48], making the outputs unstable. This is an issue that remains to be resolved, especially in the mobile communication system. We assume ΔFs as the sample rate drift and substitute it into Equ. 5 and Equ. 6. The output frequency alters as

$$f_0' = n \times (Fs + \Delta Fs) + f_2, \tag{7}$$

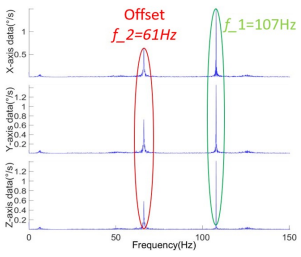


Figure 6: A sample of frequency offset.

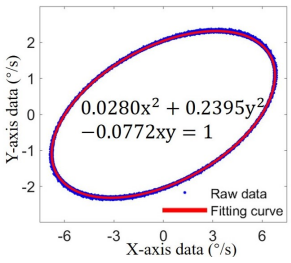


Figure 7: A scatter plot and its fitting curves.

$$R[k] = bLA \cdot sin(2\pi f_2 \frac{k}{Fs + \Delta Fs} + \Phi), \tag{8}$$

where $f_2 = f_1 - n \times \Delta Fs$ is the frequency of gyroscope readings under drift. Since f_0 is usually hundreds of times more than Fs, slight fluctuations in the sampling rate may initiate a remarkable frequency offset, as indicated in Fig. 6.

5.2 Inter-axial Characteristics

We demonstrate the offset-independent characteristics and correct the frequency offset, then further suppress motion influence on the basis of these interrelationships. Previous studies focused on the resonant data on only one axis and neglected the relation among multiple axes. We thoroughly investigate these inherent inter-axial characteristics.

Frequency synchronization. The oscillation of each axis in a gyroscope coincides. They originate from the same ultrasonic input, undergo the identical digitization process, and share the same response frequency correspondingly. The sampling rate shift, if any, is destined to happen at the same time, and accordingly, the frequency offset occurs simultaneously. Fig. 6 provides an illustration.

Fixed phase difference. Because of the synchronous resonance and the identical digitization process, the phase difference is only introduced in sensing and resonance stages. We experimentally discover that each axis oscillates at the same frequency with a fixed phase difference throughout. The scatter plot in Fig. 7 exemplifies the relationship between every two variables and supports this perspective. The curve fit an ellipse, reflecting these variables follow fixed phase difference (b). The phase characteristics of the forced vibration in Fig. (b) account for this difference. Axes differ in the natural frequency owing to production, indicating that the peak point f_N is different. When subjected to the same frequency of vibration, the ratios f/f_N in multiple axes are unequal, bringing about a difference in amplitude coefficient b and phase ϕ_1. Notably, the amplitude coefficient and phase difference keep invariable when input frequency does not change, even disturbed by motion.

5.3 Offset-independent Correction

Considering the existence of synchronous frequency and fixed phase difference, we employ a multiplier and a mean filter for correction. We multiply data in any two axes and obtain a result composed of constant bias and a second harmonic component.

Taking data from X- and Y-axes for example,

$$S_{cor}[k] = R_x[k] \times R_y[k] = \frac{1}{2}A_xA_y[cos(\Phi_x - \Phi_y) - cos(4\pi\frac{f_0'}{Fs}k + \Phi_x + \Phi_y)], \quad (9)$$

where $R_x[k]$ and $R_y[k]$ are the readings on two axes and A_x, A_y, Φ_x and Φ_y are amplitudes and phases respectively. After filtered, the harmonic component is removed, with the constant retained. The whole process is not sensitive to frequency nor offset. By aid of these relationships, we avoid the undesirable outcomes triggered by drift dexterously.

6 SYSTEM DESIGN

We propose a novel communication system that utilizes the susceptibility of gyroscopes to ultrasound. It involves combined efforts from several modules, as illustrated in Fig. 5.

6.1 Receiver Identification

Receiver identification is fundamental for communication in an enormous and dynamic IoT network. However, there are certain drawbacks associated with the use of traditional methods. Recognizing devices manually is widely used in covert channels but impractical. Meanwhile, routing protocol and address resolving demand an excessive configuration, especially in a dynamic situation. Hence, it is a conundrum to keep a balance between overheads and automation. Device fingerprint may acquit itself splendidly in this scenario. Fig. 4 reveals that different kinds of gyroscopes have various resonant frequency ranges. However, these ranges may coincide and it is difficult to further distinguish different gyroscopes of the same kind. Furthermore, we perceive the otherness of gyroscopes of the same model in resonant passband finely. *Deaf-Aid* leverages this diversity as a device fingerprint to identify receivers in a dynamic position.

From this perspective, we conduct an exploratory experiment to get the accurate resonant passband ranges at intervals of 1Hz. Conventionally, the frequencies corresponding to $\sqrt{2}/2$ of the peak value of gain coefficient are deemed starting and ending points (f_s and f_e in Fig. 3), as shown in Fig. 8. Distinctly, each gyroscope of the same model varies in the passband. Additionally, we observe each axis in one gyroscope may differ slightly. We make a comparison of each axis among gyroscopes, as diagrammed in Fig. 10. The difference of the natural frequency f_N, spawned by the production

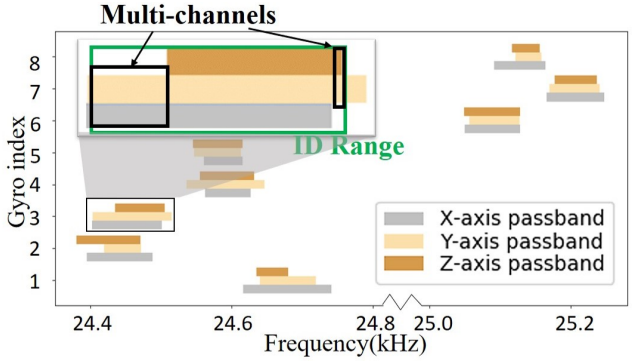


Figure 8: The bandwidth for 8 identified BMI160 chips.

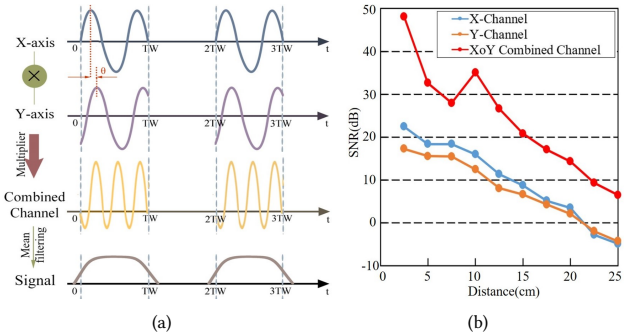


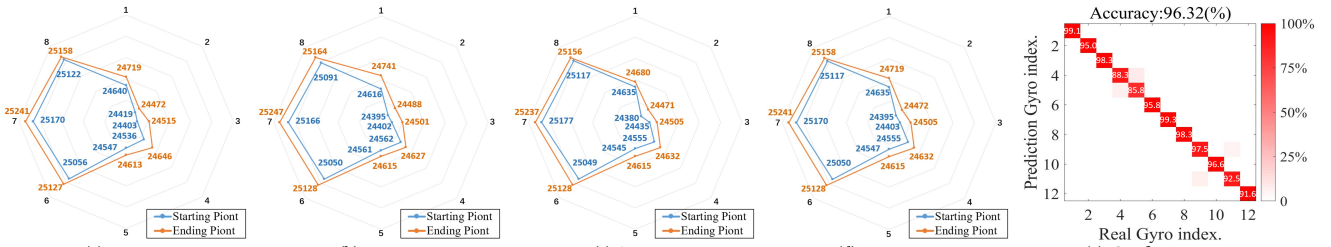
Figure 9: An illustration of (a) multiplier-based channel and (b) its performance on noise reduction.

error, accounts for it. We select the frequency range where at least two axes have a response as the ID range (the green frame in Fig. 8). It supports faster authentication than the ergodic comparisons in each axis and avoids the confusion where some gyroscopes share similar ranges on one axis. The radar chart in Fig. 10(d) confirms the validness of this fingerprint. For stable communications, the resonant frequency of a gyroscope is measured in advance and all information is known by legal users. The whole measurement process is very fast (within several minutes) and multiple devices in the same model can be measured simultaneously.

Practically, the transmitter sends an identifier composed of chirps modulated by the ID range of the target. It pushes the gyroscope to oscillate with the homologous chirps, even if there is an offset or movement disturbance. The device that receives a full identifier will be regarded as a communication target. Although utilizing the ID range reduces the dimension of features, it can still distinguish hundreds of devices. To verify the stability of this feature, we prepared an experiment a month later after the ranges were first measured. We tested 6 speakers and 12 chips of two kinds of models, including eight BMI160 chips (1-8th) and four L3GD20 chips (9-12th) with the results displayed in a confusion matrix in Fig. 10(e). It achieves an accuracy of 96.32% totally. There is a slight drop in the accuracy of the 4th and 5th chips. We note that these errors are concentrated during the test procedure via the same speaker. The poor performance of this speaker is to blame for the mistakes in identification. Even in the worst circumstances with speakers of poor frequency resolution, it still has the capability to make a distinction among tens of devices, which copes with most scenarios.

6.2 Encoding

The traditional method highlights the amplitude envelope which needs mass data to make up one bit at the sacrifice of speed. One bit should be composed of as few data as possible with a low error rate. Another issue is the signal energy fluctuation. It is not predetermined and varies along with the location and energy of sound sources. As a compromise, we make some adjustments on pulse interval encoding (PIE) [39]. We encode the data by defining different time gap widths between the rising edges of the pulses where a short interval indicates '1' and a long one implies '0', as shown in Fig. 5(b). We stipulate that only one rising edge is detected in a bit. This is conducive to the subsequent motion influence suppression.



(a) X-axis (b) Y-axis (c) Z-axis (d) ID range (e) Confusion matrix

Figure 10: Radar charts of the resonant frequency range for 8 identical BMI160 chips and confusion matrix of gyroscopes identification for 12 subjects of 2 kinds of models.

6.3 Noise Reduction

The sound wave intensity, as well as channel energy, declines as the distance increases. Discussed thoroughly in Stebler *et al.* [43], the inherent noise of a gyroscope is regarded as independent Gaussian white noise in each axis. Filtering fails due to an unstable frequency offset. In this scenario, we remove noise sagaciously based on the multiplier in Sec. 5 for signal extraction. Two axes are combined as a channel by a multiplier, with an average filter for high-frequency components and noise removal, as elaborated in Fig. 9. The combined channel has a higher signal-to-noise ratio and extends communication distance excellently.

6.4 Threshold and Decoding

It is insufficient to rely solely on empirical thresholds. The signal amplitude depends on several aspects, including communication distance, source, and resonance intensity. Inspired by the image threshold, we introduce the maximum entropy threshold method [26]. The basic idea is to find the maximum entropy and take the corresponding threshold as the final one. Concretely, for a channel with resolution r and maximum value $K \cdot r$, we decide threshold $q = k \times r$, $(k = 1, 2, \ldots, K)$ when the following entropy reaches a maximum,

$$H(q) = -\sum_{i=1}^{k} \frac{p(i \times r)}{\sum_{j=1}^{k} p(j \times r)} log \frac{p(i \times r)}{\sum_{j=1}^{k} p(j \times r)} - \sum_{i=k+1}^{K-1} \frac{p(i \times r)}{\sum_{j=k+1}^{K-1} p(j \times r)} log \frac{p(i \times r)}{\sum_{j=k+1}^{K-1} p(j \times r)}, \tag{10}$$

where $p(\cdot)$ is the probability density. Then, we decode the signals where the points with a value larger than this threshold are regarded as high level, or as low level otherwise.

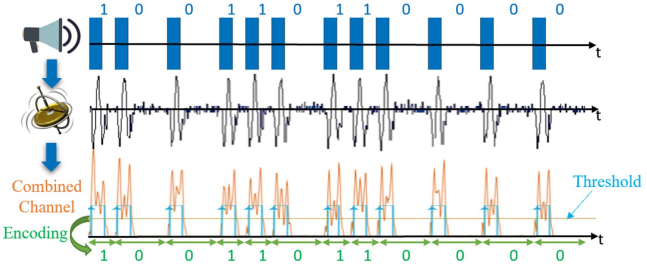


Figure 11: An example of signal transmission.

6.5 Multi-channel Support

Although resonance on each axis is dependent, it varies in the resonant frequency range. In some frequencies, only some of the axes oscillate. We choose where just two axes resonate as multi-channels (the black frames in Fig. 8). The mutual interference on the common axis can be reduced in the same way in Sec. 6.3. Thus, these channels can deliver different massages over different frequency ranges at the same time. **It provides double capacity or allows a receiver to listen to two users simultaneously.**

Consequently, we have established a gyroscope-based communication channel. A sample of signal transmission is illustrated in Fig. 11. We meet the demand for faster transmission speed with less noise, whereas it is still prone to error in movable conditions.

(a) Tx moving (b) LOS blocking (c) Rx moving.

Figure 12: Three basic kinds of motion interference.

7 MOTION INFLUENCE SUPPRESSION

Motion exerts a huge impact, especially on the gyroscope-based system. It is possible to be either disturbed by obstacle occlusion or significantly affected by the mixture of motion and resonance during communication. Fig. 12 illustrates that motion influences can be resolved into three simple forms: transmitter (Tx) moving, blocking in the line-of-sight, and receiver (Rx) moving, which can combine to form complex motion in practice. We offer an exhaustive description of the motion effect and propose solutions accordingly for robust communication in a mobile IoT network. It integrates a string of techniques, including adaptive threshold, interleaver, wavelet transform, multiplier and blind source separation.

7.1 Transmitter Motion

Transmitter motion contributes to a variation in the transmission distance. It changes the force imposed on the gyroscope, and results in signal fluctuation, including amplitude $A[k]$ and initial phase $\phi_0[k]$. The gyroscope output is rewritten as

$$R[k] = bL \cdot A[k] \cdot sin(2\pi f_1 \frac{k}{Fs} + \Phi[k]), \tag{11}$$

where, $\Phi[k] = \phi_0[k] + \phi_1 + \phi'$. Because of the signal jitters, a fixed threshold struggles and promotes the probability of error, which initiates communicating instability.

We assume the effect of distance change maintains stable in a pulse. A pulse is roughly measured in milliseconds, and the communication distance will not drastically change in such a short time. The intensity and phase changes are negligible within a pulse. Inspired by the idea of threshold window in image recognition [61], we calculate threshold in a short time (such as several bits) to achieve adaptive threshold segmentation, handling the fluctuation of the pulse. Moreover, we can normalize amplitude on the basis of these thresholds.

7.2 Line-of-Sight Blocking

Sound transmission is affected by the medium especially on LOS. The transmitter can deliberately avoid protracted obstacles. In real scenarios, it is more likely to occur disorderly, thrown in like a sudden error. The sudden error can be denoted as $SE[k]$, and the gyroscope output is rewritten as

$$R[k] = bL \cdot A \cdot sin(2\pi f_1 \frac{k}{Fs} + \Phi) + SE[k]. \tag{12}$$

We utilize interleaving technology to reduce these burst errors. Interleaving allocates the transmission bits in the time or frequency domain or both. It changes the information structure to the greatest extent without content alternation. In this way, the decoder can treat these errors as random ones, which indicates that it maximizes the dispersion of concentrated errors during channel transmission. One of the most common ways is block interleaver [5]. It writes the input sequence into a $m \times n$ matrix in the order of rows and then reads by columns. The read and write objects are swapped during reordering. The mapping function is expressed as

$$I(i) = \lceil (i-1) \bmod n \rceil + \lfloor (i-1)/n \rfloor + 1, \tag{13}$$

where $I(i)$ is the location of the ith $(i = 1, 2 \ldots, N)$ data in the original line, m and n are the number of rows and columns, respectively, $\lfloor \cdot \rfloor$ is the floor function, and $N = m \times n$ represents the interleaving length. It maximizes the dispersion of the burst errors in the process of channel transmission and effectively cuts down the errors aroused by sudden block.

7.3 Receiver Motion

Receiver motion triggers distance variation and devotes gyroscopes to produce additional readings concurrently. With distance variation solved and normalization on the basis of the adaptive threshold in Sec. 7.1, the gyroscope output is rewritten as

$$R_i[k, M_i] = b_i LA' \cdot sin(2\pi f_1 \frac{k}{Fs} + \Phi) + M_i[k], \tag{14}$$

where $M_i (i = x, y, z)$ represents the additional readings introduced by movement on the corresponding axis, and A' is the normalized amplitude. Particularly, b and ϕ_1 are immune to motion. The inter-axial characteristics are valid with a moving receiver and employed for signal recovery.

7.3.1 Motion Recovery

The initial objective of the gyroscope is to measure movements. We are supposed to eliminate the influence of motion on the signal concurrently with recovering motion. Since the resonant data is sinusoidal with a peak LbA and the frequency f_1, the accumulative error on the angle is tiny, with the maximum error $LbA/f_1\pi$. Besides, the frequency of the sinusoidal oscillations often exceeds that of motion. So it could be removed easily by wavelet transform at the cost of an acceptable loss of accuracy in angular rate measure.

Due to the random frequency offset generated by drift, this approach results in an accuracy loss in signal transmission. It is essential to separate resonant oscillation from the motion for accurate communication. We discuss it in two situations where the receiver is moving in a plane or space.

7.3.2 Signal Separation from Plane Motion

Plane motion is ubiquitous like cars, smart assistants, or cleaning robots. It affects some of axes in a gyroscope, which indicates that $M_i[k]$ in Equ. 14 do not necessarily exist synchronously. For instance, a motion is concentrated on the XoY plane, which indicates that the $M_z[k]$ is zero constantly here. Under such circumstances, these motions can count as noise. We multiply the data in X- or Y-, and Z-axes to generate a combined channel. Taking X-axis for example, it can be modeled as

$$\begin{aligned} S_{comb}[k] &= R_x[k, M_x] \times R_z[k, 0] \\ &= R_x[k, 0] \times R_z[k, 0] + M_x[k] \times R_z[k, 0]. \end{aligned} \tag{15}$$

It introduces an item $M_x[k] \times R_z[k, 0]$, where the energy of the low-frequency components, if any, is low, and the high-frequency components are removed by the mean filter, since $M_x[k]$ is often low-frequency. In this way, the plane receiver movement induces no alteration in signal transmission.

7.3.3 Signal Separation from Spatial Motion

Spatial motion is more widespread and complicated. Its complexity invalidates the multiplier-based signal extraction. We leverage the single-channel blind source separation (BSS) method with the ensemble empirical mode decomposition (EEMD) for error-free channels under spatial motion interference.

The length of intervals of rising edges between bits must fall into a definite range ruled in Sec. 6.2. Once the interval length exceeds this range, it is judged to be disturbed by motion during data transmission.

A BSS model can be represented by

$$X = AS, \tag{16}$$

where $X = [x_1[k], x_2[k], \ldots, x_N[k]]^T$ is the observation matrix, $S = [s_1[k], s_2[k], \ldots, s_M[k]]^T$ is the source matrix, and A is a $N \times M$ matrix. Typically, it requires that the number of independent observers is not less than the number of sources, that is $N \geq M$. Thanks to the encoding rules in Sec. 6.2, resonance must occur in the odd number pulse width with variance greater than the mean one. We take the subsequent high-frequency parts after wavelet transform and energy normalization, a mix of the resonant data

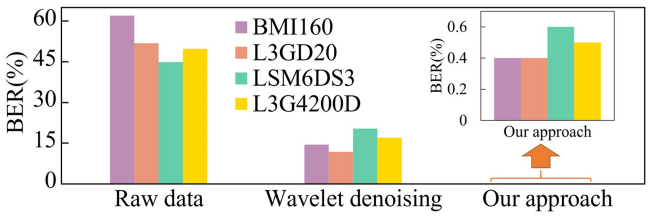


Figure 13: The performance improvement on each component of signal separation from spatial motion.

Algorithm 1: Components Reorganization Based on the Inter-axial Characteristics

Input: The n-dimension matrices $\hat{S}_1$ and $\hat{S}_2$;
Output: The resonant data $D_1[k]$ and $D_2[k]$

Initialize $A_{max} = 0$;
$B_i = [B_{i1}, B_{i2}, \ldots, B_{i\frac{n(n-1)}{2}}](i = 1, 2)$ are Power Set of $\hat{S}_i$, where $B_{ij}(j = 1, 2, \ldots \frac{n(n-1)}{2}) \subset \hat{S}_i$;
$T_{ij}[k] = \sum B_{ij}[k]$;
$L \leftarrow$ the length of $T_{ij}[k]$;
for $i \in [1, \frac{n(n-1)}{2}]$ **do**
  for $j \in [1, \frac{n(n-1)}{2}]$ **do**
    $Q = \emptyset$;
    for $m = 1 : L$ **do**
      $Q = Q \cup (T_{1i}[m], T_{2j}[m])$;
    end
    Q fits an ellipse E ;
    $\xi \leftarrow$ the mean square error of fitting;
    if $\xi < 0.01$ **then**
      $A \leftarrow$ Area of ellipse E;
      if $A > A_{max}$ **then**
        $A_{max} = A$; $i_{max} = i$; $j_{max} = j$;
      end
    end
  end
end
return $D_1[k] = P_{1i_{max}}[k]$; $D_2[k] = P_{2j_{max}}[k]$;

and remnant of motion, as the observation X, with $N = 1$ here. However, because of frequency offset and motion, there are several independent source vectors, that is $M > 2$. It is a necessity to decompose observation for $N < M$ here.

EEMD [55] is employed to decompose the single-channel mixed data to fulfill the requirement on the dimension of observation matrix. Different from FFT, EEMD manages non-stationary signal analysis. It is based on the data itself and does not require any basic function, making it more suitable for arbitrary data. It decomposes single-channel data into several intrinsic mode functions (IMFs). They constitute a n-dimensional matrix as the observation X instead. The detailed extraction has been elaborated in Huang *et al.* [21] and Mijovic *et al.* [34].

After satisfying the dimension requirement $N = n \geq M$, we utilize Fast ICA [23], a widely used solution for BSS, with a n-dimension matrix $\hat{S} = [\hat{s}_1[k], \hat{s}_2[k], \ldots, \hat{s}_n[k]]$ as a result. Nevertheless, the number of the source is unclear because of the complexity of the motion component, and there is no law on how to combine those vectors into the resonant data.

Components are reorganized on the basis of the inter-axial characteristics. On account of the fixed phase difference, resonant data on any two axes can fit a circle, or in the vast majority of cases, an ellipse. We repeat the aforementioned processes on mixed data from another axis and list all possible combinations for fitting. The one with the largest area and accredited mean square error is regarded to contain only resonant data, with detailed flow clarified in Algorithm 1. Fig. 13 reflects that our method is better than only using wavelet transform, in which the BERs are reduced to below 0.7%. It demonstrates that the BSS method and the inter-axial characteristics dominate signal recovery.

To summarize, we explicate the influence of motion and provide corresponding solutions, and as a result, prepare the system for the robustness against movement. To the best of our knowledge, it is the first work to develop the reutilization of inertial sensors with a high precision in a movable scene.

8 EVALUATION

8.1 Experimental Setup and Metrics

We build the prototype of *Deaf-Aid* using off-the-shelf devices. We conduct a comprehensive study to evaluate the accuracy and robustness of our system. These devices are fixed into brackets and the distances are adjustable, as shown in Fig. 14. These speakers play modulated signals, where the gyroscope chips' readings are collected by an Arduino and an application is developed to record gyroscope readings inside phones. The output-data-rate is set as 200Hz and the pulse width is 50ms unless otherwise stated.

Shannon channel capacity [30], a theoretically achievable upper bound, is widely used to measure the effectiveness. It is based upon the realized bit error rate, and in a binary symmetric channel, we have the channel capacity as follows,

$$C = \frac{1}{PW}[1 + BER \times log_2 BER + (1 - BER) \times log_2(1 - BER)], \quad (17)$$

where BER is the realized bit error rate and PW is the pulse width. This bound could be approached practically with proper encoding like turbo-codes [45].

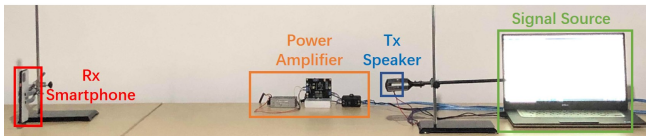


Figure 14: Experimental setup.

8.2 Universality

We take the variety of devices and environments into account to further verify the universality of *Deaf-Aid*. Here we place the speaker and gyroscope at a distance of 15cm in three different locations including a large seminar room, a small office, and a crowded laboratory. We test on six speakers of three kinds (including JBL 750T [19], Samsung Galaxy S8 [15], and HIVI-SS1II [20]), whose supply power is limited within 5W, and 32 gyroscopes of four models

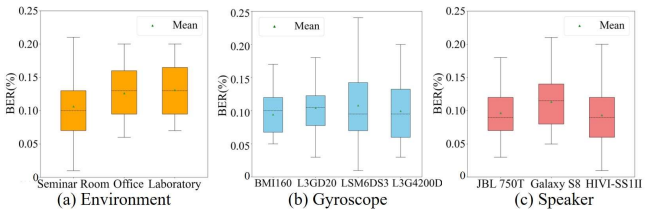
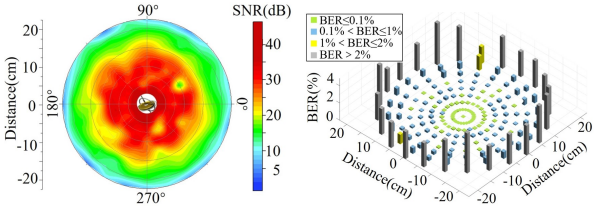
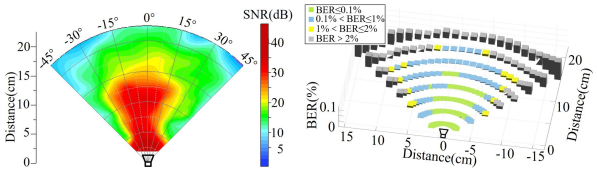


Figure 15: The impact of environment and devices.



(a) SNR (b) BER

Figure 16: Performance centered on the gyroscope.



(a) SNR (b) BER

Figure 17: Performance centered on the speaker.

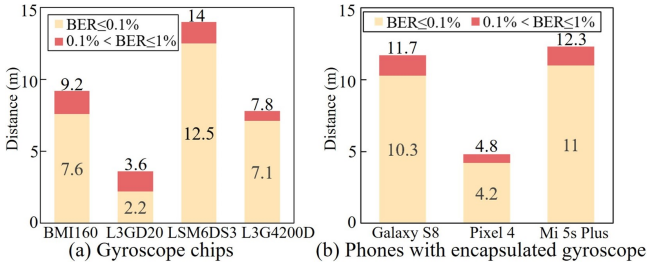
(including BMI160, L3GD20, LSM6DS3, and L3G4200D). The distribution is presented by box-plots in Fig. 15. In these situations, our system performs diversely but satisfactorily, with BER lower than 0.25% comprehensively. It guarantees a stable communication quality among numerous devices with little deformation due to the ambient environment.

8.3 Orientation and Distance

Placement limitation is a crucial issue in covert channels. Taking a 5W JBL750T speaker and a BMI160 gyroscope chip as an example, we examine the resilience under multiple layouts to further explain that there is no restriction on the layout of devices. Concretely, we rotate the speaker around the fixed gyroscope. The performance in the XoY plane is illustrated in Fig. 16 and that of other planes are similar. It reflects the placement of gyroscopes makes no difference. Conversely, we rotate the gyroscope around the fixed speaker. Fig. 17 illustrates the effectiveness in the range of a 22.5° opening angle of speakers, with a BER of 0.1% at 15cm and 1% at 20cm. It is practical for users tend to turn towards the objective, and slight direction deviation is tolerable. Moreover, we carry with arbitrary layout and draw similar results.

The above results about the communication range are obtained from a power-limited speaker, whose power setting is only 5W. We adopt such a setting with the consideration that some IoT devices are equipped with such a power-limited speaker for the purpose of low power consumption.

For those devices with less strict requirements on power consumption, the communication distance can dramatically increase. For example, we raise the power of the speaker to 30W, which is also very common in existing commodity speakers. **The distance can be extended up to 14m**, as indicated in Fig. 18(a). In general, different gyroscopes have different resonance peaks b_{max}, and different communication distances correspondingly. Nevertheless, even the L3GD20 chip, which is with the worst performance among our gyroscopes, can support a communication distance of 3.6m.



(a) Gyroscope chips (b) Phones with encapsulated gyroscope

Figure 18: Communication distance.

Such a communication range is sufficient to cover most application scenarios. It can be concluded that *Deaf-Aid* is capable of supporting error-free and remote transmission with scarce constraints on placement in real life scenarios.

8.4 Transmission Capacity

Transmission speed is conditional on output-data-rate (ODR) and pulse width (PW). In this evaluation, a pair of the speaker (JBL GTO 750T, 5W) and gyroscope (BMI160) are placed 15cm apart in the seminar room to judge the trend of transmission capability with different ODR and PW.

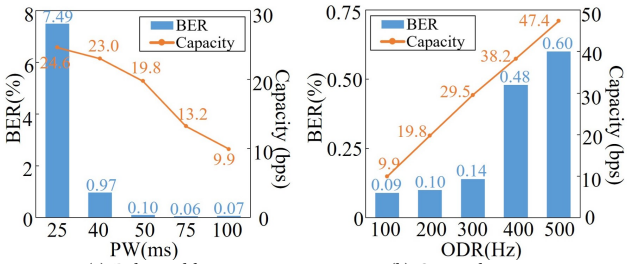
Pulse width. We appraise our system with an adjustable pulse width in the range of 25ms and 100ms when the ODR in gyroscope is defaulted to 200Hz, a widely used rate in mobile devices. The product PW×ODR decides the amount of data contained in a bit. A shorter PW understandably means fewer data to form a bit and possibly cause more errors in this issue. In Fig. 19(a), the results vividly demonstrate that BER maintains below 1% when PW is longer than 40ms and 0.1% when PW is longer than 50ms.

Output-data-rate. Similarly, we repeat the experiment where ODR varies evenly between 100Hz and 500Hz at 100Hz intervals. Fairly, we maintain the product PW×ODR at 0.01. As illustrated in Fig. 19(b), channel capacity ascends with the incline of ODR, up to 47.4bps. Although there is a slight increase in BER, it remains a low level within 0.6%.

Generally speaking, *Deaf-Aid* is competent for the different requirements of transmission speed and tolerance of error flexibly in various occasions.

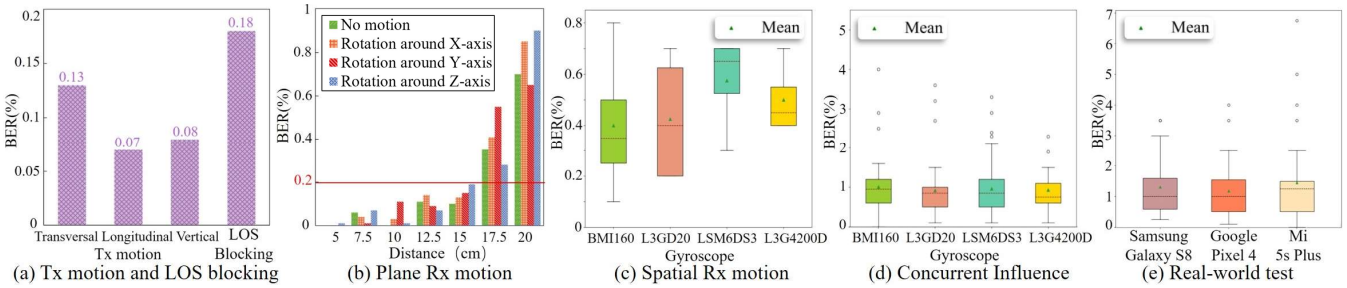
8.5 Multi-channel

By the aid of multi-channel, communication capacity can double, or one receiver can get information simultaneously from two transmitters in different scenarios. The third chip in Fig. 8 is exemplified



(a) Pulse width (b) Output-bit-rate

Figure 19: Performance under different conditions.



(a) Tx motion and LOS blocking (b) Plane Rx motion (c) Spatial Rx motion (d) Concurrent Influence (e) Real-world test

Figure 20: Performance of *Deaf-Aid* against motion interference.

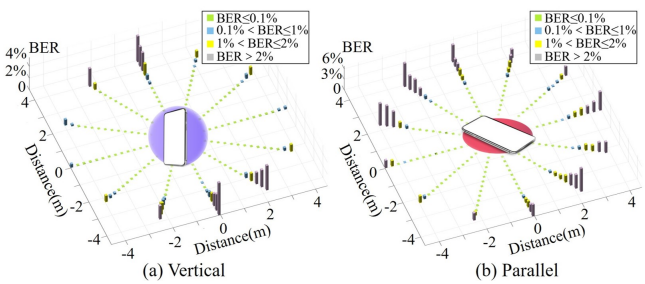
Table 1: Validity of simultaneous multi-channel

Mode	Channels	BER(%)	Capacity(bps)
Case1	XoY	0.23	38.79
	YoZ	0.39	
Case2	XoY	0.27	19.46
	YoZ	0.17	19.64

to bear out the feasibility of multi-channel, where XoY channel works at 24.41kHz and YoZ channel works at 25.5kHz. We test in two cases. In case 1, a speaker delivers different massages on these two channels simultaneously. In case 2, two speakers each deliver on one of them respectively, with the performance attached to Tab. 1. In both cases, it succeeds at the expense of a slight accuracy loss. ***Deaf-Aid* supports simultaneous communication on multiple channels, even from two transmitters.**

8.6 Smartphone Prototype

To verify the feasibility of *Deaf-Aid* working on those devices with encapsulated gyroscopes, we build the smartphone prototype using three kinds of phones, including Samsung Galaxy S8, Google Pixel 4, and Mi 5s Plus. All those devices are with encapsulated gyroscopes, corresponding to LSM6DSL, BMI160, and ICG-20660/L, respectively. We measure the transmission distance using a 30W speaker. Fig. 18(b) reflects that our system is able to communicate with those phones up to 12.3m away, as their screen is set vertical to the ground. The BMI160 encapsulated in Pixel 4 shows the shortest communication distance, 4.4m. But it is still satisfactory in many scenarios, e.g., the indoor environment.



(a) Vertical (b) Parallel

Figure 21: Performance centered on a Pixel 4.

We use a Pixel 4 to further validate the flexibility of *Deaf-Aid* in terms of layout. We rotate the speaker around the fixed phone, which is vertical and parallel respectively, with graphical representations of results in Fig. 21. The communication distance fluctuates between 3.1m and 4.8m with BER less than 1%. This enables a smartphone to retrieve massages sent within three meters accurately via *Deaf-Aid*, no matter in which orientation it is. This demonstrates that *Deaf-Aid* is capable of establishing communication among realistic devices in the wild, merely a tiny penalty of slightly shortening the communication range.

8.7 Motion Influence

Plenty of deliberate motion interference is involved for a better understanding of the robustness of *Deaf-Aid* against the movement. We bind a 5W JBL speaker, an obstacle, and gyroscopes to a manipulator respectively. They move under the control of the program. The experimental distance is set within 15cm by default. Moreover, we recruit 22 participants, who arm with 30W speakers and three kinds of phones with encapsulated gyroscopes for further confirmation on real-world scenes.

Rx motion and LOS blocking. We manipulate a speaker into moving in three simple directions (transverse, longitude, and verticality), while a BMI160 chip is fixed. Then we manipulate an obstacle into moving randomly in LOS between the fixed speaker and gyroscope. As graphically shown in Fig. 20(a), BER is around 0.1% when the speaker moves and is always below 0.2% even under obstacle disturbance. There is no doubt that it can settle the impact of transmitter motion and LOS blocking smoothly.

Plane Tx motion. Then we fix the speaker and rotate a BMI160 chip around its axes with the comparison at different distances illustrated in Fig. 20(b). This system maintains a low BER. It is less than 0.2% at a distance of 15cm and rises to 1% as the distance increase to 20cm. Thereby, the plane motion of gyroscopes has little effect on the stability of our system.

Spatial Tx motion. Here, gyroscopes move in space irregularly within 15cm from a fixed speaker. This evaluation involves four types of gyroscopes and each type contains 8 chips. Repetitive experiments are conducted on these chips where the manipulator repeats the same trajectory. It has a maximum error of 0.8% with all averages lower than 0.7% in Fig. 20(c). We have prepared it for the robustness against the fundamental movement.

Concurrent influence. We ask 22 volunteers to send information with a speaker in hand where 32 gyroscope chips move in

Table 2: Comparison with previous work

System	Basic	Speed	Accuracy	Receiver Identification	Placement	Distance	Motion Robustness
Ripple [41]	Vibra-motor to Accelerator	200 bps	BER<1.7%	Manually	Fixed on a plane	6 inches	Not
Ripple II [40]	Vibra-motor to Microphone	30 kbps	SNR>15db	Manually	Physical contact	Touch based	Just to tiny vibration
BitWhisper [17]	Heat emission to thermal sensor	1-8 bits per hour	Not evaluate	Manually	Fixed position	40cm	Not allow move
Dhwani [36]	Speaker to Microphone	2.4 kbps	Accuracy>95%	Manually	No limitation	10cm	Yes
Deaf-Aid	Speaker to Gyroscope	47 bps	BER<0.6%	Automatic	No limitation	14m	Yes

space irregularly under the same conditions as above. As shown in Fig. 20(d), our system performs well under the multiple concurrent motion interference. It maintains the mean of BER below 1%. Although there exists off-group points data, the peak is lower than 6%. One explanation for those outliers is that volunteers accidentally deflect the orientation of the speaker away from receivers.

Real-world motion. In order to evaluate the robustness of the prototypes, we organize volunteers in pairs. In each pair, one volunteer holds a 30W speaker and the other carries a smartphone. Both of them freely move within a range of 2 meters and fiddle with the devices. We evaluate on three kinds of smartphones and find that the mean of BER is below 1% and the peak is lower than 7% during the entire experiment, as the result shown in Fig. 20(e). This indicates that *Deaf-Aid* facilitates a robust channel among mobilizable IoT devices in a real-world implementation.

In conclusion, *Deaf-Aid* shows immense potential as a communication bridge even under various motion disturbances in a complicated IoT network.

8.8 Comparison with Previous Systems

Covert channels take advantage of physical phenomenon for data transmission among adjacent devices. We select some typical cases for comparison, listed in Tab. 2. Communication through vibration, for example *Ripple* [40] and *Ripple II* [41], is good at speed but weak at fixed position and poor motion robustness. *BitWhisper* [17] delivers messages further but slowly on a covert channel using thermal manipulations. The speaker-to-microphone channel is exploited by *Dhwani* [36]. However, with the purpose of recording human voices, the microphones on IoT devices are more likely to filter out 8kHz [1]. In this case, the *Dhwani* like approaches require peripherals to utilize ultrasound for stealthy communication, such as a high-quality microphone and a sound card with a high sampling rate. Otherwise, people nearby will be disturbed. *Deaf-Aid* has no such issues instead, not to mention that *Deaf-Aid* also has other advantages, such as multi-channel communication and automatic receiver identification.

In summary, *Deaf-Aid* enables IoT devices to identify and chat with their neighbors remotely, infallibly, and liberally. It provides an alternative and complementary communication channel to current IoT devices.

9 DISCUSSION

9.1 Implementation Consideration

Communication distance and capacity will soar along with technology. A better speaker, with a wider spectrum of responses or larger power, extends the communication range. It is reported that ultrasound is capable to affect gyroscopes 37m away [42]. This indicates a great potential of *Deaf-Aid* in more scenarios. On the other hand, increasing the sampling rate would result in a higher transmission rate. Therefore, *Deaf-Aid* would contentiously improve in the transmission rate upon the emergence of new hardware. For example, the gyroscopes MPU6050 [25] and BMI160 [8] that are used in our experiments support 1kHz and 3.2kHz sampling rates, respectively. If we adopt a gyroscope with an over 10kHz sampling rate, *Deaf-Aid* can raise the transmission rate to thousands of bps by a conservative estimate. We will obtain a more efficient system as new hardware emerges.

Signal clipping means that the sensing mass produces voltages exceeding the input range of its amplifier, and distort the signal. For instance, it occurs on communication via an L3GD20 chip within 5cm experimentally. Even so, rising edges are still recognizable. Clipping introduces little additional error statistically. We will further analyze the sensitivity of gyroscopes under different conditions, for example, sound pressure levels, to find the optimal device setting.

Power consumption is another issue that deserved discussions. Currently, in the audio components of IoT devices, the power is almost consumed by the speaker, relied on the volume rather than frequency. Its rated power and maximum power are fixed after the manufacture. In general, the consumption of a commodity speaker is designed and constricted within an acceptable range for an IoT device. For the transmitters, it is just the consumption of speakers and we have limited the power of speakers in our experiments, with the consideration on the low power consumption of IoT devices in some cases (see Sec. 8.3). In addition, if the transmitter is a mobile phone, the power consumption would not be a big issue due to the aid of the high-capacity battery and power bank. *Deaf-Aid* prepares itself for occasions with both limited and sufficient power supply.

More IoT devices and platforms will be supported in the foreseeable future. On the basis of proven resonance phenomenon in 3D mouse, screwdriver, VR device, iPhone [48], drone [42] and remote control model car [47], our system can be applied in a broader range

of devices including those above. This could be an essential step to expand application fields, thus leading a more comprehensive IoT network based on *Deaf-Aid*.

9.2 Security

The current resonant frequency band is relatively narrow, determined by the inherent structure of gyroscopes. In this case, some sophisticated communication techniques, such as FDM and OFDM, are not applicable. However, we exploit the potential of the narrow bands from a security perspective.

Jamming: It is difficult for jammers to find out the appropriate band of a gyroscope, narrower than 50Hz usually. Without sufficient knowledge, an attacker has to jam in a broadband spectrum. This method demands professional acoustic devices and it can be detected easily. A practicable means of avoiding malicious jamming is timely jamming detection. It is easy for *Deaf-Aid*, as only a microphone is required.

Eavesdropping: *Deaf-Aid* can prevent replay attacks even if the private key is leaked. Benefited from gyroscopes' narrow band-pass width, intended non-informative ultrasound signal could broadcast at the nearby frequency to confuse attackers but receivers are impervious to those noises with the help of multiplier-based signal extraction. Furthermore, users can utilize multi-channel with signals on one channel and deceptive data on the other. Prior information is an absolute necessity for eavesdroppers, such as the communication frequency band, which is difficult to pick out the right one from camouflage.

10 RELATED WORK

Privacy is recorded by inertial sensors. A malicious attacker can easily obtain inertial sensors data inside mobile platforms without access permission, for keystroke inference [9, 29, 35, 37, 50, 58], device identification [12, 60] and speech recognition [2, 3, 18, 33]. The *Gyrophone* [33] like approaches leverages a gyroscope as an eavesdropper to recognize speeches, mostly lower than 1kHz. Their intention is to eavesdrop on the context of human conversation via vibration. Different from *Gyrophone*, *Deaf-Aid* benefits from the resonance of gyroscopes and is aiming at transferring modulated information from a speaker to a gyroscope.

Gyroscope is vulnerable to acoustic injection attacks. It has been demonstrated that resonance of gyroscopes could be triggered by acoustic signals [10, 11, 47]. An adversary can impose on outputs of gyroscopes, bringing about control systems error. A DoS attack was conducted to incapacitate drones [42]. Tu *et al.* [48] realized a black box switching attack to push victim gyroscope to produce expected outputs.

Covert channels have attracted great interest. They take advantage of physical phenomena, such as heat [4, 17, 56, 57], light [32, 44], electromagnetic leakage [16], and ultrasound [31]. Inertial sensors have become candidates [6, 13, 40, 41]. Nevertheless, these methods demand physical contact, specialized equipment or artificial assistance, none of which is needed in *Deaf-Aid*.

11 CONCLUSION

In this paper, we leverage the stealthy speaker-to-gyroscope channel for protocol-independent mobile IoT communication. We first probe the relationship among axes in a gyroscope under resonance and triumph over adversity, such as frequency offset and multi-channel communication support. As an innovation, the diversity of resonant frequency range among gyroscopes is employed as fingerprint for automatic receiver identification. The motion influence suppression and corresponding mobile communication have been delicately designed. Our system, *Deaf-Aid*, reaches up to 47bps with a low BER even under motion interference. It could act as a stepping-stone for an everything-related IoT network.

ACKNOWLEDGMENTS

This paper is partially supported by National Natural Science Foundation of China under grant 61872285, 61772236, and 61972348, the Fundamental Research Funds for the Central Universities, Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Research Institute of Cyberspace Governance in Zhejiang University, Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (Grant No. 2018R01005), Zhejiang Key R&D Plan (Grant No. 2019C03133).

REFERENCES

[1] H. Abdullah, W. Garcia, C. Peeters, P. Traynor, K. R. B. Butler, and J. Wilson. 2019. Practical Hidden Voice Attacks against Speech and Speaker Recognition Systems. In *Proceedings of the Network and Distributed System Security Symposium*.
[2] S. Anand and N. Saxena. 2018. Speechless: Analyzing the Threat to Speech Privacy from Smartphone Motion Sensors. In *Proceedings of the IEEE Symposium on Security and Privacy*.
[3] Z. Ba, T. Zheng, X. Zhang, Z. Qin, B. Li, X. Liu, and K. Ren. 2020. Learning-based Practical Smartphone Eavesdropping with Built-in Accelerometer. In *Proceedings of the Network and Distributed System Security Symposium*.
[4] D. B. Bartolini, P. Miedl, and L. Thiele. 2016. On the capacity of thermal covert channels in multicores. In *Proceedings of the Eleventh European Conference on Computer Systems*.
[5] C. Berrou and A. Glavieux. 1996. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Transactions on Communications* 44, 10 (1996), 1261–1271.
[6] K. Block, S. Narain, and G. Noubir. 2017. An Autonomic and Permissionless Android Covert Channel. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*.
[7] C. Bormann, A. P. Castellani, and Z. Shelby. 2012. CoAP: An Application Protocol for Billions of Tiny Internet Nodes. *IEEE Internet Computing* 16, 2 (2012), 62–67.
[8] Bosch. 2018. BMI160 Datasheet. https://www.bosch-sensortec.com/products/motion-sensors/imus/bmi160.html.
[9] L. Cai and H. Chen. 2012. On the Practicality of Motion Based Keystroke Inference Attack. In *Trust and Trustworthy Computing - 5th International Conference*.
[10] R. N. Dean, S. T. Castro, G. T. Flowers, G. Roth, A. Ahmed, A. S. Hodel, B. E. Grantham, D. A. Bittle, and J. P. Brunsch. 2011. A Characterization of the Performance of a MEMS Gyroscope in Acoustically Harsh Environments. *IEEE Trans. Industrial Electronics* 58, 7 (2011), 2591–2596.
[11] R. N. Dean, G. T. Flowers, A. S. Hodel, G. Roth, S. Castro, R. Zhou, A. Moreira, A. Ahmed, R. Rifki, B. E. Grantham, D. Bittle, and J. Brunsch. 2007. On the Degradation of MEMS Gyroscope Performance in the Presence of High Power Acoustic Noise. In *IEEE International Symposium on Industrial Electronics*.
[12] S. Dey, N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi. 2014. AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable. In *Network and Distributed System Security Symposium*.
[13] B. Farshteindiker, N. Hasidim, A. Grosz, and Y. Oren. 2016. How to Phone Home with Someone Else's Phone: Information Exfiltration Using Intentional Sound Noise on Gyroscopic Sensors. In *10th USENIX Workshop on Offensive Technologies*.
[14] GSMArena. 2017. Mobile Release Date, Price, Specs, Features, Review, News. https://www.gsmarena.com.
[15] GSMArena. 2017. Samsung Galaxy S8. https://www.gsmarena.com/samsung_galaxy_s8-8161.php.
[16] M. Guri, G. Kedma, A. Kachlon, and Y. Elovici. 2014. AirHopper: Bridging the air-gap between isolated networks and mobile phones using radio frequencies. In *9th International Conference on Malicious and Unwanted Software*.
[17] M. Guri, M. Monitz, Y. Mirski, and Y. Elovici. 2015. BitWhisper: Covert Signaling Channel between Air-Gapped Computers Using Thermal Manipulations. In *IEEE 28th Computer Security Foundations Symposium*.

[18] J. Han, A. J. Chung, and P. Tague. 2017. PitchIn: Eavesdropping via Intelligible Speech Reconstruction Using Non-acoustic Sensor Fusion. In *16th ACM/IEEE International Conference on Information Processing in Sensor Networks*.

[19] Harman. 2019. JBL STADIUM GTO750T. https://www.onlinecarstereo.com/CarAudio/p_51143_JBL_STADIUMGTO750T.aspx.

[20] HiVi. 2013. SS1II. http://www.hivi.com/products/detail.aspx?pid=100000443464649.

[21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London* 454, 1971 (1998), 903–995.

[22] U. Hunkeler, H. L. Truong, and A. Stanford-Clark. 2008. MQTT-S — A publish/subscribe protocol for Wireless Sensor Networks. In *3rd International Conference on Communication Systems Software and Middleware and Workshops*.

[23] A. Hyvarinen. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10, 3 (1999), 626–634.

[24] IDC. 2019. Worldwide Internet of Things Forecast, 2019–2023. https://www.idc.com/getdoc.jsp?containerId=US45373120.

[25] Invensense. 2015. MPU-6050 Datasheet. https://www.invensense.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf.

[26] J. Kapur, P. Sahoo, and A. Wong. 1980. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing* 29 (1980), 273–285.

[27] J. Lee, Y. Su, and C. Shen. 2007. A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi. In *33rd Annual Conference of the IEEE Industrial Electronics Society*.

[28] K. Lee, H. Wang, and H. Weatherspoon. 2014. PHY Covert Channels: Can you see the Idles?. In *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation*.

[29] X. Liu, Z. Zhou, W. Diao, Z. Li, and K. Zhang. 2015. When Good Becomes Evil: Keystroke Inference with Smartwatch. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*.

[30] C. T. M. 2017. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience.

[31] A. Madhavapeddy, R. Sharp, D. J. Scott, and A. Tse. 2005. Audio networking: the forgotten wireless technology. *IEEE Pervasive Computing* 4, 3 (2005), 55–60.

[32] A. Maiti and M. Jadliwala. 2019. Light Ears: Information Leakage via Smart Lights. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 98:1–98:27.

[33] Y. Michalevsky, D. Boneh, and G. Nakibly. 2014. Gyrophone: Recognizing Speech from Gyroscope Signals. In *Proceedings of the 23rd USENIX Security Symposium*.

[34] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel. 2010. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE transactions on biomedical engineering* 57, 9 (2010), 2188–2196.

[35] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury. 2012. Tapprints: your finger taps have fingerprints. In *The 10th International Conference on Mobile Systems, Applications, and Services*.

[36] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan. 2013. Dhwani: secure peer-to-peer acoustic NFC. *ACM SIGCOMM Computer Communication Review* 43, 4 (2013), 63–74.

[37] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang. 2012. Accessory: password inference using accelerometers on smartphones. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*.

[38] J. Ploennigs, U. Ryssel, and K. Kabitzsch. 2010. Performance analysis of the EnOcean wireless sensor network protocol. In *IEEE 15th Conference on Emerging Technologies Factory Automation*.

[39] P. R. Prucnal and P. A. Perrier. 1988. Optical self-routing in a self-clocked photonic switch using pulse-interval encoding. In *Fourteenth European Conference on Optical Communication*, Vol. 1. 259–263.

[40] N. Roy and R. R. Choudhury. 2016. Ripple II: Faster Communication through Physical Vibration. In *13th USENIX Symposium on Networked Systems Design and Implementation*.

[41] N. Roy, M. Gowda, and R. R. Choudhury. 2015. Ripple: Communicating through Physical Vibration. In *12th USENIX Symposium on Networked Systems Design and Implementation*.

[42] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim. 2015. Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors. In *Proceedings of 24th USENIX Security Symposium*.

[43] Y. Stebler, S. Guerrier, J. Skaloud, and M. Victoria-Feser. 2012. A framework for inertial sensor calibration using complex stochastic error models. In *Proceedings of the IEEE/ION Position, Location and Navigation Symposium*.

[44] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu. 2019. Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems. arXiv:2006.11946

[45] P. Sweeney. 2002. *Error Control Coding: From Theory to Practice*. John Wiley Sons, Inc.

[46] W. T. Thomson. 1981. *Theory of vibration with applications*. Prentice Hall.

[47] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu. 2017. WALNUT: Waging Doubt on the Integrity of MEMS Accelerometers with Acoustic Injection Attacks. In *IEEE European Symposium on Security and Privacy*.

[48] Y. Tu, Z. Lin, I. Lee, and X. Hei. 2018. Injected and Delivered: Fabricating Implicit Control over Actuation Systems by Spoofing Inertial Sensors. In *Proceedings of 27th USENIX Security Symposium*.

[49] S. Vinoski. 2006. Advanced Message Queuing Protocol. *IEEE Internet Computing* 10, 6 (2006), 87–89.

[50] C. Wang, X. Guo, Y. Wang, Y. Chen, and B. Liu. 2016. Friend or Foe?: Your Wearable Devices Reveal Your Personal PIN. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*.

[51] Z. Wang, W. Zhu, J. Miao, H. Zhu, C. Chao, and O. K. Tan. 2005. Micromachined thick film piezoelectric ultrasonic transducer array. *Sensors Actuators A Physical* 130-131 (2005), 485–490.

[52] A. Wheeler. 2007. Commercial Applications of Wireless Sensor Networks Using ZigBee. *IEEE Communications Magazine* 45, 4 (2007), 70–77.

[53] Wikipedia. 2020. Gyroscope. https://en.wikipedia.org/wiki/Gyroscope.

[54] Wikipedia. 2020. Hi-Fi. https://en.wikipedia.org/wiki/High_fidelity.

[55] Z. Wu and N. E. Huang. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis* 1, 1 (2009), 1–41.

[56] Z. Wu, Z. Xu, and H. Wang. 2012. Whispers in the Hyper-space: High-speed Covert Channel Attacks in the Cloud. In *Proceedings of the 21th USENIX Security Symposium*.

[57] Z. Wu, Z. Xu, and H. Wang. 2015. Whispers in the Hyper-Space: High-Bandwidth and Reliable Covert Channel Attacks Inside the Cloud. *IEEE/ACM Transactions on Networking* 23, 2 (2015), 603–615.

[58] Z. Xu, K. Bai, and S. Zhu. 2012. TapLogger: inferring user inputs on smartphone touchscreens using on-board motion sensors. In *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*.

[59] C. Yum, Y. Beun, S. Kang, Y. Lee, and J. Song. 2007. Methods to use 6LoWPAN in IPv4 network. In *The 9th International Conference on Advanced Communication Technology*.

[60] J. Zhang, A. R. Beresford, and I. Sheret. 2019. SensorID: Sensor Calibration Fingerprinting for Smartphones. In *IEEE Symposium on Security and Privacy*.

[61] D. Zhou and W. Cheng. 2008. Image denoising with an optimal threshold and neighbouring window. *Pattern Recognition Letter* 29, 11 (2008), 1694–1697.

[62] Y. Zou, J. Zhu, X. Wang, and L. Hanzo. 2016. A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends. *Proc. IEEE* (2016).