IACR Transactions on Cryptographic Hardware and Embedded Systems
ISSN 2569-2925, Vol. 2024, No. 2, pp. 654–676. DOI:10.46586/tches.v2024.i2.654-676

Laser-Based Command Injection Attacks on Voice-Controlled Microphone Arrays

Hetian Shi[1], Yi He[1], Qing Wang[3], Jianwei Zhuge[1,2], Qi Li[1] and Xin Liu[4]

[1] Tsinghua University, Beijing, China
shiht18@tsinghua.org.cn, heyi21@mails.tsinghua.edu.cn
zhugejw@tsinghua.edu.cn, qli01@tsinghua.edu.cn
[2] Zhongguancun Laboratory, Beijing, China
[3] Huawei Cloud Co., Ltd., Beijing, China
sereinnwang@163.com
[4] Lanzhou University, Lanzhou, Qinghai, China
bird@lzu.edu.cn

Abstract. Voice-controlled (VC) systems, such as mobile phones and smart speakers, enable users to operate smart devices through voice commands. Previous works (e.g., LightCommands) show that attackers can trigger VC systems to respond to various audio commands by injecting light signals. However, LightCommands only discusses attacks on devices with a single microphone, while new devices typically use microphone arrays with sensor fusion technology for better capturing sound from different distances. By replicating LightCommands's experiments on the new devices, we find that simply extending the light scope (just as they do) to overlap multiple microphone apertures is inadequate to wake up the device with sensor fusion. Adapting LightCommands's approach to microphone arrays is challenging due to their requirement for multiple sound amplifiers, and each amplifier requires an independent power driver with unique settings. The number of additional devices increases with the microphone aperture count, significantly increasing the complexity of implementing and deploying the attack equipment. With a growing number of devices adopting sensor fusion to distinguish the sound location, it is essential to propose new approaches to adapting the light injection attacks to these new devices. To address these problems, we propose a lightweight microphone array laser injection solution called LCMA (Laser Commands for Microphone Array), which can use a single laser controller to manipulate multiple laser points and simultaneously target all the apertures of a microphone array and input light waves at different frequencies. Our key design is to propose a new PWM (Pulse Width Modulation) based control signal algorithm that can be implemented on a single MCU and directly control multiple lasers via different PWM output channels. Moreover, LCMA can be remotely configured via BLE (Bluetooth Low Energy). These features allow our solution to be deployed on a drone to covertly attack the targets hidden inside the building. Using LCMA, we successfully attack 29 devices. The experiment results show that LCMA is robust on the newest devices such as the iPhone 15, and the control panel of the Tesla Model Y.

Keywords: laser command injection · voice-controlled systems · photoacoustic effect · pulse-width modulation · laser transmitters array · electrostatic effect

1 Introduction

Voice-controlled (VC) systems are used ubiquitously in various devices such as smart home appliances, mobile devices, and Intelligent Connected Vehicles(ICVs) in our daily

Received: 2023-10-15 Accepted: 2023-12-15 Published: 2024-03-12

lives. Smart speakers like Google Home Assistant, Amazon Echo, and Apple HomePod demonstrate the growing trend of controlling smart devices with voice commands. However, this trend also introduces new attack surfaces where adversaries use audio command injection to control smart home appliances and electric vehicles (EVs).

Efforts to attack VC systems generally fall into two categories: network-based and sensor-based attacks [RHRC17, MZL20, DLZZ14a, YLZ+20]. Network-based attacks, targeting software vulnerabilities, are often resolved through firmware updates [MBM+18, AWS+19]. In contrast, sensor-based physical attacks manipulate physical vibrations in microphones using ultrasound [ZYJ+17], laser [SKKK17], or electromagnetic waves [DAY22], and may necessitate combined software and hardware modifications for mitigation.

LightCommands proposes the first laser-based audio injection attack for VC systems, which can convert audio commands into light signals to trigger the VC devices. However, it mainly targets single microphone devices and overlooks multi-microphone VC devices with sensor fusion technology and non-MEMs (Micro-Electromechanical Systems) microphones, including ECM (Electret Condenser Microphones) and Piezoelectric types. With the growing prevalence of complex multi-microphone systems, conventional strategies like enlarging the laser beam are becoming obsolete. Nor is it possible to use LightCommands's approach to target different apertures of microphone arrays with multiple laser beams. This is because LightCommands uses amplitude modulation (AM) for signal conversion, which requires cumbersome equipment such as audio amplifiers and power drivers. Their method requires setting unique light frequencies for different apertures, which is impractical and time-consuming due to the need for extensive manual configuration of modulation parameters. These drawbacks make LightCommands inefficient in multi-microphone scenarios [SCR+20], especially for attacking EV control panels.

In this paper, we introduce LCMA (Laser Commands for Microphone Array), an advanced laser-based audio injection attack that extends the scope of LightCommands to multi-microphone sensor fusion situations and non-MEMs devices. LCMA uses a Pulse Width Modulation (PWM) algorithm and a laser transmitter array to digitize audio signals, which can effectively mitigate the impact of environmental noise. This innovation eliminates the need for frequent adjustments post-setup. A key contribution of LCMA is to overcome the challenges of compromising sensor fusion in Voice-controlled (VC) systems by directing different laser signals with specific phase differences to each microphone in the array. This method successfully bypasses VC system defenses that rely on sound source location detection. LCMA takes advantage of the ubiquity and efficiency of PWM modules in MCUs like the STM32F407, a cost-effective solution at as low as $11.3, for converting audio signals into precise laser commands, offering a stark cost advantage over traditional AM modulation systems. This adaptability makes LCMA not only a theoretical model but also a viable, cost-effective practical solution that can even be deployed via drones, ushering in a new era of vulnerability exploration for VC systems.

We have conducted extensive testing of LCMA on 29 different models of devices, with 23 of them not previously examined by earlier studies, especially the three devices equipped with non-MEMs microphone(ECM, piezoelectric microphone). Remarkably, the results show that all of these devices are universally vulnerable to our LCMA approach, which effectively bypasses existing defenses, including those provided by LightCommands and subsequent research[XZJX21]. LCMA's novel laser array design can concentrate laser energy on individual microphones to defeat traditional light-barrier-based defenses. In addition, the laser's internal reflection within the audio channels can even bypass L-shaped channel defenses via light infiltration. Consequently, we propose new strategies to robustly defend against advanced laser signal injection attacks, and provide mathematical analysis and experimental validation for them.

The contributions of LCMA are as follows:

- We introduce a new approach that combines unipolar PWM modulation with a laser

array to enable extensive attacks on underexplored microphone array devices. LCMA significantly expands the scope of laser injection attacks by providing a simplicity, scalability (e.g., supporting devices with multiple microphones or non-MEM systems), and cost-effectiveness solution compared to previous methods like LightCommands.

- We conduct a thorough evaluation of LCMA using 29 different models of VC devices. We find that even the latest VC systems, such as the Tesla Model Y's control panel & iPhone 15, are still vulnerable to laser attacks. We delve into the fundamental physical reasons behind laser attacks.
- We demonstrate LCMA's ability on effectively compromising VC devices with existing defense measures, highlighting both the method's advanced capabilities and the limitations of current defensive strategies. In turn, this encourages us to propose new, more robust defense strategies tailored to better protect against the sophisticated threats posed by laser-based attacks.

2 Background

2.1 Voice-Controlled System

Voice-controlled (VC) devices increasingly become more popular for their ability to interpret and respond to voice commands in natural language, offering a user-friendly interface [CBR+13]. These systems are designed to promptly respond to spoken commands, such as "Turn off the light," where a voice-controlled device would immediately execute the command. VC devices primarily differ in their microphone structures, which can be categorized into single microphone wake-up and multi-microphone wake-up devices. This distinction is crucial in understanding their wake-up mechanisms and response patterns, as illustrated in Figure 1.

(a) Single Microphone Wake-Up Scenario (b) Multi-Microphone Wake-Up Scenario

Figure 1: Difference Between Single Microphone and Multi-Microphone Wake-Up Scenarios

The wake-up requirements for VC devices equipped with a microphone array structure differ from those with only one or two microphones. As depicted in Figure 1(a) and 1(b), the red circle signifies the microphone that captures user's voice commands. In Figure 1(a), the left microphone captures the commands, while in Figure 1(b), the left three adjacent microphones receive the signal simultaneously. Typically, VC devices designed for single-microphone wake-up usually incorporate just one or two microphones, any of which can successfully trigger the device upon capturing the user's voice command.

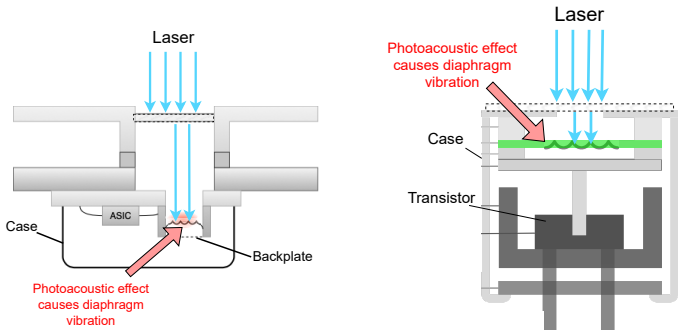
In contrast, VC devices with a microphone array structure utilize multiple microphones to receive user commands. To achieve a successful wake-up of the device, the voice command must be captured by multiple microphones within the array. In mainstream smart speakers, more than half of all microphones in the array are necessary to effectively respond to a command.

Microphone Array

Integrating microphone arrays into VC devices is a popular industry practice, due to the enhanced sound capture and voice recognition capabilities they offer. These arrays enable precise voice command recognition, crucial for applications like in-car systems, and improve noise cancellation by focusing on the sound source and reducing background noise.[KMR12, BW01]. They are available in three main configurations: linear arrays, which are cost-effective but offer limited noise reduction; planar arrays, which provide a 360-degree pickup on a plane and are suited for devices like smart speakers; and 3-D arrays, which offer the best omnidirectional sound capture but at a higher cost, used in premium products like Apple's Homepod I and Vendor-A's SoundJoy.

Typical Microphone Types

According to the Electrostatic effect, there are several kinds of microphones: MEMs microphone, moving-coil microphone, electret condenser microphone (ECM), and some other types. In this paper, we mainly focus on MEMs and ECM. Figure 2 shows these two microphones structure.



(a) MEMs (b) ECM

Figure 2: Two Types of Microphones

MEMs microphones, as illustrated in Figure 2(a), feature a condenser structure design. In this configuration, a powerful laser beam penetrates the microphone's dust net, reaching the internal diaphragm polarized by an Application-Specific Integrated Circuit (ASIC). The sudden alteration in light energy serves as the trigger for diaphragm oscillations.

ECM microphones, portrayed in Figure 2(b), also utilize a condenser structure design. Here, a potent laser light traverses the microphone's dust net to reach the pre-charged diaphragm of the ECM, generating photoacoustic effects that cause diaphragm oscillations.

Both types of microphones utilize a condenser structure, relying on a capacitor formed between a stationary backplate and a movable diaphragm. Sound pressure variations cause the diaphragm to move, altering the capacitor's charge and converting sound waves into electrical signals. Their operational principles, rooted in the Electrostatic effect, make them susceptible to laser-based audio injection attacks.

2.2 PWM: Voice Signal to Laser Conversion

Pulse Width Modulation (PWM) is a central technique in LCMA for converting analog voice signals into digital format suitable for laser control. This process involves modulating the width of digital signal pulses to reflect the amplitude of analog signals [Gol92].

As shown in Figure 3, the duty cycle of PWM, defined as Duty Cycle $= \frac{\text{Time ON}}{\text{Total Period}} \times 100\%$, dictates the duration the signal remains high within the total cycle time. Accurate voice signal sampling, adhering to the Nyquist theorem [Nyq], is essential for capturing the full information of audio. Post-sampling, the voice signal is converted into a PWM

signal, where variations in pulse width are proportional to audio amplitude changes. Such conversion techniques, particularly the use of unipolar PWM [San93, OAV04].

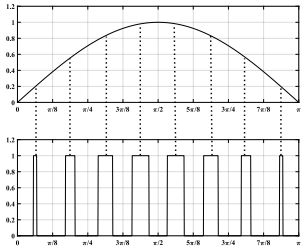


Figure 3: Pulse Width Modulation Theory

In LCMA, the resultant PWM signal precisely controls laser emissions, ensuring VC devices interpret the laser signal as a legitimate voice command. The system includes a Bluetooth receiver and a PWM-configured development board, digitizing the voice signal and encoding its amplitude variations into the PWM duty cycle. This method enables LCMA to replicate complex voice commands through controlled laser intensity variations, showcasing its advanced capabilities in VC device interaction.

2.3 Hardware Specifications and Advantages for LCMA

LCMA leverages the advanced capabilities of STM32 micro-controller units (MCU), such as the STM32F103 and STM32F407 series. These chips are chosen for their high count of PWM outputs, critical for LCMA's complex signal processing requirements. The STM32F103, for example, offers up to ten PWM outputs, while the STM32F407 provides up to 16 [harb, hara]. This abundance of PWM ports allows LCMA to address sensor fusion scenarios effectively, a crucial advantage over previous methods like LightCommands which are limited by fewer Digital-to-Analog Converter (DAC) ports and suffer from quantization errors [harc]. This hardware setup positions LCMA as a robust and versatile solution in the realm of audio injection attacks for voice-controlled systems.

2.4 Laser-Based Attacks

Lasers, valued for their coherence, monochromatic properties, and high brightness, have been widely utilized in cryptography and fault injection. They have the capacity to target critical components like Physically Unclonable Functions(PUFs) and encryption chips to disrupt security protocols [TLG+15, BJC15]. In autonomous driving systems, the vulnerability of sensors like LiDAR to laser interference raises significant safety concerns [YXL16, SCCM20, SKKK17].

LightCommands method developed by Sugawara et al. for laser-based commands injection in VC devices. It highlights the challenges faced when dealing with devices that use sensor fusion technology, which typically involves multiple microphones to improve sound capture. The method's limitations include its inability to simultaneously trigger all microphones with a single laser spot and its sensitivity to environmental factors affecting the SNR (Signal-to-Noise Ratio). Additionally, the high cost of the necessary equipments, particularly the laser driver, is emphasized, indicating a need for a more cost-effective solution for attacking VC devices. The breakdown of the costs for a single setup, totaling over $348, underscores the financial aspect of these limitations [SCR+20].

3 Related Works

In the domain of sensor-based attack scenarios on voice-controlled devices, previous researches have identified multiple ways of injection attack. We categorize them into the

following three groups:

Audible Command Injection This class of attacks involves injecting voice commands that are either genuinely spoken or software-generated into voice-controlled (VC) systems. Malicious entities have engineered applications capable of producing artificial voice commands to compromise VC devices without requiring authentication [DLZZ14b]. Although these attacks are inherently detectable due to their audible characteristics, research has evolved towards camouflaging voice commands as signals that evade human detection yet remain interpretable by speech recognition systems [VZSS15, WM21]. It is, however, pertinent to note that such modified signals may retain detectability to the human ear, which poses a risk of discovery [SM17].

Inaudible Command Injection This strategy seeks to obscure voice commands from human perception entirely, using high-frequency sounds beyond the range of human hearing but within the capture capabilities of standard microphones [RHRC17]. Recent advancements have enabled the transmission of entirely inaudible commands to VC systems by exploiting the non-linearities of microphone circuits to modulate signals on ultrasonic carriers [ZYJ+17]. Despite limitations like short effective range and potential for partial audible leakage, innovations such as signal decomposition, and the use of loudspeaker arrays have improved the reach and effectiveness of these attacks [RHRC17].

Laser Injection This innovative approach deploys modulated laser beams to inject commands into MEMs microphone-equipped devices. Compared to audible and inaudible methods, laser-based techniques have the advantage of being undetectable by human sense of hearing and can target devices from a distance through transparent media. The LightCommands method is a prominent instance, although it encounters challenges like restricted efficacy against devices with multiple microphones, sensitivity to the parameters of the attack environment, and the elevated costs associated with setups [SCR+20]. To overcome these challenges and provide a cost-effective solution for multi-microphone systems, we introduce the Laser-based Command Modulation Attack (LCMA), an innovative approach that enhances the feasibility and practicality of audio command injection into various voice-controlled systems.

4 LCMA Overview

4.1 Motivation

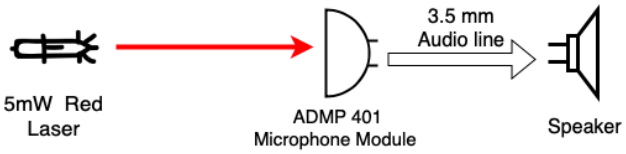


Figure 4: Experiment: Laser Pointer Shining on the ADMP401 MEMs Module

In our study, we address the challenge of attacking devices with microphone arrays, which requires simultaneous control of multiple lasers. This inspires us to employ PWM for modulating voice signals, based on the hypothesis that variations in laser intensity can excite the microphones to generate an acoustic response.

To test this hypothesis, we conduct an experiment using an ADMP401 microphone module connected to a speaker to simulate the audio pick-up function of a VC device, as shown in Figure 4.

Experiment-1: We test this with a 5mW red laser and an ADMP401 microphone module connected to a speaker, simulating a VC device. By oscillating an obstruction

between the laser and microphone at about 3Hz, we observe a distinct "clicking" sound from the speaker.

The phenomenon supports our hypothesis that MEMs microphones could “translate” light intensity variations. We then explored the modulation of voice signals into changes in light intensity using unipolar PWM signals, which are ideal for this digital transmission and can be easily generated by MCUs.

Experiment-2: A smartphone served as an audio source, playing the music “Narco”, is connected to a signal generator(model: UTG1005A) through a 3.5mm audio jack, which in turn is connected to a 1.6W, $\lambda = 450nm$, Laserland laser transmitter aimed at the ADMP401 microphone module. We configure the UTG1005A in PWM mode with the following parameters: PWM frequency $(f) = 20kHz$, output signal peak-to-peak voltage $(V_{pp}) = 5V$, bias voltage $(V_{offset}) = 2.5V$, and duty cycle $(D_{Duty}) = 50\%$.

The speaker successfully plays the rhythmic music “Narco”, matching the smartphone playback. This demonstrates the viability of using PWM for audio-to-laser conversation. More details and audios from this experiment are available on our website https://github.com/Moriartysherry/Silent-Attack.

LCMA effectively resolves sensor fusion challenges in voice-controlled devices by leveraging the greater number of PWM output ports in MCUs compared to DACs. This, combined with the chips’ timer functionalities, enables LCMA to precisely deliver distinct signals to each microphone in an array, a critical requirement for overcoming sensor fusion complexities.

4.2 LCMA Design

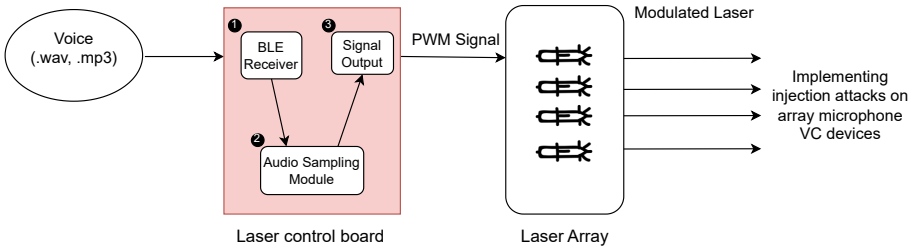


Figure 5: System Architecture of LCMA

The architecture of LCMA is shown in Figure 5 which contains three main steps:

1. **Audio Recording and Transmission** Initially, audio, in formats like .wav or .mp3, is initially recorded and transmitted to the laser control board’s receiver module via Bluetooth.
2. **Audio to PWM Signal** Secondly, we use the laser control broad to convert audio into PWM signal, which consists of three key modules:
   - (1)Bluetooth (BLE) Module: This module receives voice commands in .wav format from a PC or mobile device via Bluetooth.
   - (2)Audio Sampling Module: This component converts analog audio files into digital samples using an Analog-to-Digital Converter (ADC).
   - (3)Signal Output Module: The digitized audio is modulated into Pulse Width Modulation (PWM) signals, which are used to control the laser array.
3. **Modulated Laser Beams** Thirdly, the control signals from the board are then converted into modulated laser beams, which are precisely targeted at the VC device’s microphone.

The attack process successfully manipulates a VC device by directing modulated laser beams at its microphone, causing it to execute commands as if they are regular audio inputs.

For the issues of laser transmission, as shown in Figure 6, the analog audio signal is then quantized into a unipolar PWM signal's duty cycle, transforming it into a digital representation. This PWM signal transmits 'digitalized command signals' to VC devices via lasers, ensuring consistency across different commands. It preserves essential parameters like signal-to-noise ratio, and information content of the voice signals. LCMA efficiently generates phase offsets tailored for the transmission of multiple laser signals, facilitating precise and synchronized signal injection into the microphone array of the targeted VC devices.

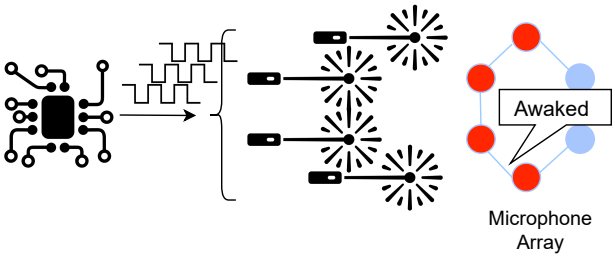


Figure 6: LCMA for Simultaneous Laser Transmitter Control

The system design includes two transmission modes in the STM32 development board: aiming and attacking. The aiming state operates at low light intensity through set PWM signal duty cycle into 5%, allowing the operator to fine-tune the laser emitter array's position and alignment in relation to the target device's microphone array. An adjustable laser stand aids in this precise alignment process.

In the attacking phase, LCMA reverts to normal power and employs real human voice recordings as the signal source, effectively circumventing voice-print detection systems. The attacker's pre-prepared voice signal is channeled through a 3.5mm audio cable and a Bluetooth signal receiver to the STM32 development board's input port.

4.3 LCMA Threat Model

To launch a laser attack with LCMA, we assume that the attackers have a direct line of sight to the targeted VC devices so that the laser beams can be aimed at the microphones. This sight may not necessarily be a horizontal straight path, as tools like drones could be used to achieve the required angle. The attacker's goal is to remotely inject commands to VC devices via lasers, without producing any detectable sound, aiming for precise control and responses from the devices. To perform laser-based injection, attackers may employ different tools to remotely align the laser with the device's microphones. These tools can include gears for precise adjustment of each laser beam's position or drones equipped with gimbal stabilizers for precise laser aiming at the microphones. They can also monitor device responses, such as LED lights and audible cues, to confirm if their attack is successful.

We assume attackers can grasp the necessary characteristics of the target device, like microphone array layouts, to fine-tune their attack strategy. This assumption is reasonable because attackers can identify the types of target devices based on their appearance and purchase the same device. For devices with voice-print detection, we assume they can obtain real voices from the victim or use other ways such as voice forgery techniques to mimic the victim's voice.

4.4 How can PWM solve the sensor fusion problem?

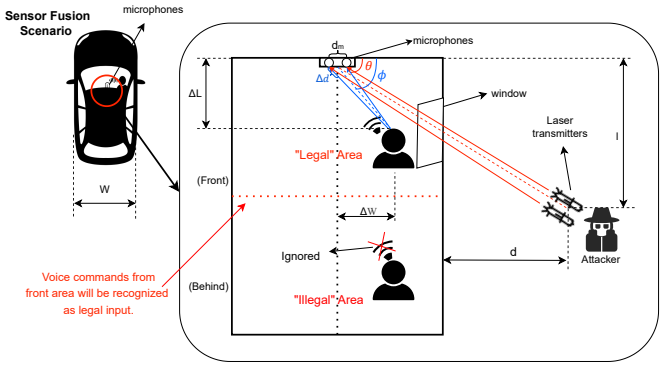


Figure 7: Sensor Fusion Scenario

Sensor fusion typically involves the integration of multiple microphones to enhance sound localization and recognition capabilities in VC devices [Aar03, dVIV+17]. For instance, in-car VC systems utilize a multi-microphone array to discern if commands originate from a designated, "Legal" area, such as responding only to specific commands from the front area, as depicted in Figure 7. Our research has found that VC systems determine the source location of a sound by comparing the time difference between signals received by two microphones. To exploit this, we can deceive the VC system by injecting two laser beams with the corresponding phase difference δ.

$$\theta = arcsin(\frac{\sqrt{(H-h)^2+l^2}}{\sqrt{(\frac{W}{2}+d)^2+(H-h)^2+l^2}}) \tag{1}$$

$$\delta = 2\pi f_{PWM} d_m [\frac{cos(arcsin(\frac{\sqrt{(\Delta H)^2+(\Delta L)^2}}{\sqrt{(\Delta W)^2+(\Delta H)^2+(\Delta L)^2}}))}{v_{sound}} - \frac{cos(arcsin\theta)}{c_{light}}] \tag{2}$$

Equation 1 and Equation 2 calculate the laser incident angle θ and the corresponding phase difference δ respectively. It uses the height difference between the laser source and microphones $(H-h)$, the lateral distance to the microphone (l), the width of the vehicle (W), the distance from the vehicle to the laser source (d), the PWM frequency (f_{PWM}), the distance between two microphones (d_m), the height, lateral, and width distances between microphones and the supposed audio source (ΔH, ΔL, and ΔW), the sound velocity (v_{sound}), and the speed of light (c_{light}).

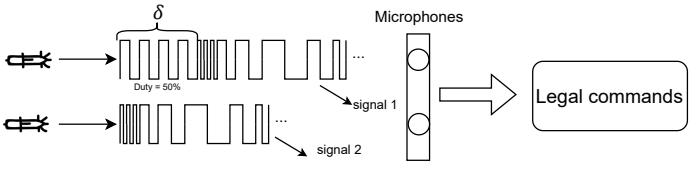


Figure 8: Solution for Sensor Fusion Issues

Utilizing auxiliary equipment, attackers can derive spatial parameters as shown in Figure 7, facilitating the calculation of the laser incident angle θ in eq 1. The essence of

LCMA lies in adjusting the laser signal's phase difference δ in eq 2, to emulate the natural time difference observed in VC system. This technique aligns the injected commands across the microphone array, effectively mimicking authentic audio signals and deceiving VC devices. Attackers can create a phase difference δ in the laser signals (Figure 8), spoofing the VC system into recognizing the laser as a legitimate sound source.

5 Experiments

In this section, we provide a detailed description of our experimental setups, procedures, and results. Our experiments are designed to evaluate the efficacy of LCMA in various attack scenarios targeting different voice-controlled systems. We have meticulously analyzed the impact of multi-dimensional environmental factors on the effectiveness of LCMA. This includes exploring how variables such as the thickness and color of transparent obstacles, like window glass or the glass used in Intelligent Connected Vehicles (ICVs), and the specific layout of microphones, especially those utilizing an L-shaped configuration, influence our system's ability to penetrate defenses. Additionally, we have included a feasibility analysis of these environmental factors to enhance our understanding and develop more effective defense strategies against LCMA. For a comprehensive view of our experiments, including videos demonstrating LCMA's application across different VC systems, visit our project website: https://github.com/Moriartysherry/Silent-Attack.

Ethics Consideration All our works are performed on our private devices, ensuring no impact on other users. Our tests comply with the security bounty programs of the respective vendors. We have disclosed our findings to all relevant vendors. These attacks have been acknowledged by them, assisting in mitigating potential threats.

5.1 Experiment Results

Our study evaluates 29 prominent VC device models, as listed in Table 1. All devices are successfully compromised, including 22 equipped with microphone arrays. Notably, four of these could be attacked using the method described in [SCR+20], which does not require activating multiple microphones simultaneously. While the remaining 18 devices need to activate multiple microphones at the same time for a successful attack, a capability beyond the scope of LightCommands. LCMA, with its unique laser array configuration, effectively overcomes this limitation, offering broader attack coverage.

Table 1: Table of VC Devices Tested by LCMA

Device Type	Vendor Name	Device Model
Smart Speakers(8)	Google, Apple, Amazon, Vendor-A, Xiaomi, Tmall	Google Home mini, Apple HomePod I, Vendor-A Smart Voice Speaker, Amazon Echo Dot 2nd gen, Tmall Genie Square Candy, Xiaomi Smart Speaker Pro, Vendor-A SoundX, Vendor-A Sound Joy
Mobile Devices & Bluetooth Headsets(10)	Apple, Vendor-A, Samsung, Xiaomi	iPhone 6, iPhone 8, iPhone 15, Vendor-A Mate 40, Vendor-A P40, Vendor-A P50, AirPods 2, Samsung Galaxy Buds live, Vendor-A Bluetooth Headset FreeBuds 4E, Xiaomi Bluetooth Headset Buds 3
Electric Vehicle (EV)(4)	Weltmeister, ARCFOX, Vendor-A,Tesla	Weltmeister EX5, ARCFOX Alpha S, AITO M5, Tesla Model Y
Smart Screen(3)	Vendor-A, Xiaomi	Vendor-A Hicar Smart Screen, Xiaomi Mijia Smart Rearview Mirror, Vendor-A Smart Screen(TV),
Others(4)	Philips, SUITU	Multi-party conference System, Little Bee Speaker(ECM), In-car Kettle(ECM), Piezo Crystal Microphone Module

The expanded experimental scope of LCMA notably includes additional laser attack scenarios on Tesla vehicles and conference systems, as well as on non-MEMs microphone devices, beyond traditional targets like smartphones and smart home devices. The laser

array and PWM signals effectively address sensor fusion challenges in the phase difference δ setups, enabling direct command injections into Tesla vehicles through windows, rather than the need to use a phone app as mentioned in LightCommands. Furthermore, LCMA's concentrated energy output allows for the injection of authentic voice signals into non-MEMs microphone devices, overcoming higher triggering thresholds of this type of microphone, which represents a significant advancement over the linearly varied frequency sine waves used in LightCommands experiments, offering the first practical implications in real-world attack scenarios.

In our experiments, we employ a methodical approach exemplified by our setup with a Vendor-A SoundX smart speaker. After procuring the unit, we establish its functionality and carefully position it to optimize laser targeting. This meticulous setup process is reflective of our broader experimental methodology across various devices.

Table 2: Number of Microphones Required for Successful Laser Injection in VC Devices

Microphone	Aperture	Representative Devices	Lasers
MEMs	6	Apple HomePod, Vendor-A Smart Voice Speaker, Xiaomi Smart Speaker Pro, Vendor-A Soundx	≥4
	4	Tmall Genie Square Candy, AirPods 2, Samsung Galaxy Buds live, Vendor-A Bluetooth Headset, Xiaomi Bluetooth Headset, AITO, Tesla Model Y	All
	2	Vimax Auto, ARCFOX Car, Vendor-A Hicar Smart Screen, Xiaomi Mijia Smart Rearview Mirror, Vendor-A Smart Screen(TV),	
	≤3	Vendor-A Sound Joy, iPhone 6, iPhone 8, iPhone 15,Vendor-A P40,Vendor-A Mate 40, Vendor-A P50, Multi-party conference System, Amazon Echo Dot 2nd gen, Google Home mini	1
ECM	1	Little Bee Speaker(ECM), In-car Kettle(ECM)	
Piezo Crystal	1	Piezo Crystal Microphones	

Our attack, detailed in Table 2, requires over 400mW power for effective device compromise. Interestingly, this high power necessity is also noted in three devices previously analyzed in LightCommands works. This implies post-vulnerability disclosure adjustments by vendors to reinforce defenses against attacks. Table 2 further delineates the specifics of our laser attack, including the microphone types in each device and the number of lasers required, underscoring the varying complexities and LCMA's adaptability across different device types.

5.2 Case Studies for Different Attacking Targets

In this section, we present three case studies for different attack targets to better demonstrate the effectiveness of the LCMA approach and its coverage of a wider range of attack scenarios.

Case Study 1: Microphone Array Parameters Adjustment

In this case study, we evaluate the efficacy of LCMA by attacking a VC device with a microphone array. A key challenge is aligning multiple lasers with the microphones using optical aids and a custom laser transmitter's mount. Additionally, we adjust the phase delays δ for each channel based on estimated laser incident angles as discussed in Section 4.4. This alignment is crucial for ensuring the VC device recognizing the injected commands as legitimate human speech. The laser's output is set to aiming power, producing a faint spot that the attacker could precisely target the microphone with by

rotating gears on the mount. For scenarios involving sensor fusion, it is only necessary to input the estimated angle θ into the control GUI, with no need to adjust other parameters such as PWM sampling rate, signal amplitude, and so on.

Due to the significant speed difference between light and sound, the value of $\frac{\cos(\arcsin\theta)}{c_{\text{light}}}$ is approximately zero. This implies that the arrangements of laser transmitters have minimal impact on the signal parameters received by the microphone. Therefore, the lasers can only be arranged in a random staggered formation in space without interfering with each other.

In the following experiments, we conduct tests on a VC device(Vendor-A SoundX) equipped with a microphones array. Each of the six lasers ($\lambda = 450$nm) of the LCMA device is precisely aligned with a microphone. The experiments involve activating different numbers of lasers, ranging from all six to just one, while simultaneously injecting three distinct commands into the VC device. The commands and their injection details are described in Table 3. Each command is injected three times, and the process will be halted upon successful device response to prevent disruptions.

Table 3: Voice Command Recording Details

	Play music	Turn up the lamp	What is the weather like today
Duration(s)	3.3	4.0	4.2
Start from(s)	0.7	0.6	0.8

Success in waking up VC devices is ascertained by their response to laser-induced 'wake up' commands, characterized by the activation of audible alerts and visual indicators. Command injection is considered successful when the devices execute actions that are congruent with the laser-modulated commands, demonstrating accurate command recognition and execution by the VC devices.

Table 4 describes the relationship between the number of laser beams and the effects of LCMA. When five or more out of the six microphones are illuminated by lasers, LCMA could consistently wake up VC devices and successfully inject commands. However, when the number of illuminated microphones drops below five, for example, in cases where only four or three microphones are targeted, the effects of LCMA require more detailed discussion. The spatial arrangement of the illuminated microphones also plays a crucial role in the effects of the attack. When the illuminated microphones are distributed, as shown in Figure 9, at least one of the illuminated microphones has neither of the two adjacent microphones illuminated by lasers. LCMA could only wake up the VC devices without successfully injecting commands. Conversely, if the illuminated microphones are adjacent, it is necessary to inject commands into at least four microphones simultaneously to achieve command injection effect into VC devices.

Table 4: Distributed VS Adjacent

#microphone	**adjacent**	**awakened**	**command injection**
≥ 5	-	✓	✓
4	yes	✓	✓
4	no	✓	×
3	yes	✓	×
3	no	✓	×
< 3	-	×	×

Case Study 2: Attacking Non-MEMs Devices

The application of MEMs microphones in smart devices predominates over other microphone types, primarily due to their high degree of integration. Nonetheless, the

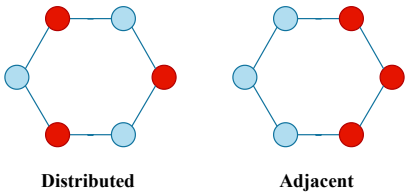


Figure 9: The Number of Apertures Required for Microphone Array Attacks

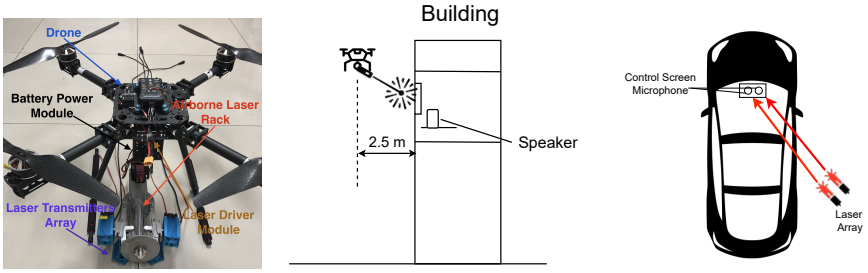
ability of our approach to inject commands into non-MEMs devices would significantly extend the reach and impact of LCMA. In order to validate the potential of our approach to also target non-MEMs microphones, we conduct tests on devices equipped with ECM and piezoelectric microphones (Piezoelectric Vibration Sensor Modules).

For the ECM microphone experiments, we select TLT-0501MZ car-mounted voice heating kettle as the attack target. We determine the success of command injection by observing the kettle's response after injecting laser commands. As for the piezoelectric crystal, we utilize the type of sensor module produced by Telesky, which consists of a piezoelectric sensor and a signal amplifier. To assess the impact of our approach on piezoelectric microphones, we connect the sensor using a 3.5mm headphone cable, play the output signal through a speaker, and determine whether our approach has any effects on the piezoelectric microphone by comparing the heard sound with the original audio.

In our test with the TLT-0501MZ car-mounted electric kettle, we successfully inject the wake-up command 'XiaoLi, XiaoLi, boil mode' using our method, leading the kettle to respond and begin boiling. Additionally, we transmit Wiz Khalifa's 'See You Again' to the piezoelectric microphone using LCMA, and could clearly hear the melody from the speaker, demonstrating our approach's effects in attacking piezoelectric microphones.

Unlike LightCommands, our experiments cover a diverse range of non-MEMs microphones, demonstrating our approach's effects in injecting both voice commands and music. These findings confirm LCMA's capability to target a wider variety of microphone technologies. All demos mentioned above can be accessed on https://github.com/Moriartysherry/Silent-Attack.

Case Study 3: Remote Attack Scenario



(a) Drone installed LCMA (b) Drone Scenario (c) Tesla In-car Scenario

Figure 10: Remote Attack Scenarios of LCMA

Our experiments about attacks in remote scenarios demonstrate that LCMA is suitable for a wider range of attack scenarios, but other laser-based attack methods do not have this capability. As shown in Figure 10(a), we mount the laser attack equipments on a drone and conduct laser attacks on indoor VC devices from an outdoor location. The schematic diagram is illustrated in Figure 10(b). While the attack range of LCMA is relatively close compared to LightCommands, the use of a drone allows us to shorten the attack distance.

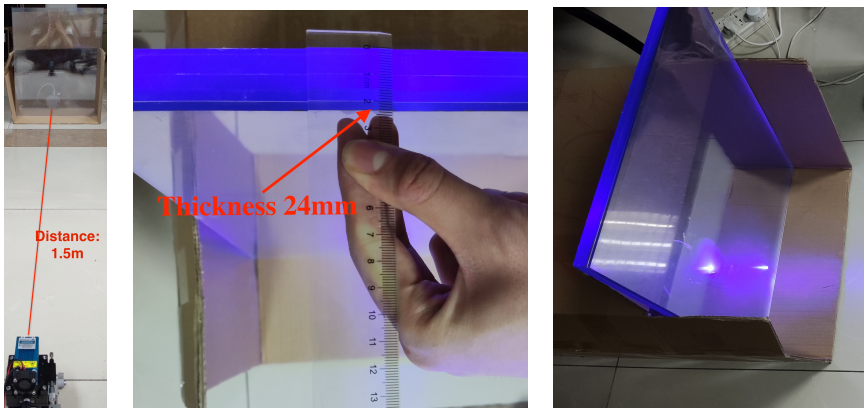
Notably, the drone-assisted alignment allows the laser array to be effectively positioned at an optimal distance of approximately 2.5 meters, facilitating a successful attack.

Furthermore, LCMA also successfully conducts a laser attack on a Tesla Model Y from outside the vehicle towards the in-car microphones equipped with advanced sensor fusion technology, leading to the successful opening of the car window. Unlike previous attacks that involve sending laser commands to a smartphone with a vehicle control app installed [SCR+20], our attack scenario is direct laser injection through the window into the in-car microphones[1], as shown in Figure 10(c).

5.3 Feasibility for LCMA

In the section, we conduct a comprehensive feasibility analysis focusing on the multi-dimensional aspects of the environment surrounding the target device. This analysis encompasses three key environmental features: the thickness of transparent media, the color of light filters, and the L-shaped structure of sound paths. Each of these factors plays a significant role in determining the susceptibility of devices to LCMA and, therefore, is critical in formulating robust defenses. Our study also demonstrates the feasibility of using infrared lasers for attacks.

Laser Penetration Efficacy Across Different PVC Thicknesses



(a) Front (b) Vertical View (c) End View

Figure 11: Command Injection Across PVC Plates with Different Thicknesses

In our comprehensive study on the impact of material thickness on laser attack efficacy, we focus on the penetration capabilities of laser beams through different thicknesses of Polyvinyl Chloride Glass (PVC) plates, representative of common window materials. As depicted in Figure 11, we conduct a series of experiments with a laser of 450 nm wavelength and 400 mW average power at a distance of 1.5 meters, simulating real-life scenarios of attacks through windows.

Our findings, detailed in Table 5, reveal a remarkable ability of LCMA to penetrate PVC of varying thicknesses, up to 23.5mm. This result is particularly significant as it surpasses the standard thickness of most commercial building windows, indicating the method's high potential in real-world settings. The data demonstrates not only the raw penetration power of the laser but also its effective use in command injection, challenging the notion of safety provided by physical barriers and calling for more advanced protective measures in VC device security.

Impact of Filter Color on LCMA Efficacy

[1] potentially located in the middle of the front seats or above the side windows.

Table 5: Command Injection Results to Different Thicknesses of PVC Plates

thickness/mm	awakened	command injection
2.5	✓	✓
3.0	✓	✓
4.0	✓	✓
5.0	✓	✓
9.0	✓	✓
11.5	✓	✓
14.0	✓	✓
18.0	✓	✓
21.0	✓	✓
23.5	✓	✓

In investigating the impact role of filter color in LCMA's effects, we explore how varying hues affect laser light absorption and transmission, thereby influencing the success of laser injections. This examination aims to evaluate the success of laser attacks when traversing colored plexiglass. We employ polymethyl methacrylate (PMMA) plexiglass plates of various colors, with a thickness of 2.5 mm. The experiments utilize a laser featuring a 450 nm wavelength and an average power output of 400 mW, positioned at a 2-meter distance. As detailed in Table 6 (indicated by '*', signifying that the success rate is not 100%), our results reveal a strong correlation between the attack effects and the colors of filters.

Our analysis reveals that for 450nm (blue) laser light, the success of penetration is inversely related to the 'B' (blue) component in the RGB makeup of the filter: lower 'B' values correlating with higher penetration rates, as shown in Table 6. This suggests that choosing a filter with minimal to zero maximum RGB components effectively blocks LCMA attacks. Furthermore, the use of a thin PMMA plate in our experiments demonstrates its potential as an effective countermeasure. This study underscores the significance of filter color selection in enhancing defenses against laser-based security threats, highlighting the potential of color-based defense strategies to mitigate the risks posed by sophisticated attacks like LCMA.

Table 6: Command Injection Results to Different Colors of PMMA Plates

glass color	Hex/HTML	awakened	command injection
dark red	#7C2230	×	×
orange red	#FF4200	×	×
orange yellow	#FFA930	×	×
coffee	#853C10	×	×
tawny	#988022	×	×
green	#3E6E3A	×	×
yellow	#FAE600	✓	✓*
purple	#781761	✓	✓
dark blue	#121563	✓	✓
sky blue	#6FD2E4	✓	✓
blue	#0047BE	✓	✓
red	#EA0447	✓	✓

Effectiveness of L-Shaped Microphone Structure in Mitigating LCMA

In our testing experiments, we find that some phone manufacturers, like Vendor-A, designed their main microphones in an L-shaped structure, as illustrated in Figure 12. This design, made feasible by the placement of the pickup port underneath MEMs microphone chips, enables sound to navigate turns that light cannot. Combined with a narrow and elongated channel, this design prevents direct light from hitting the MEMs microphone diaphragm without affecting sound collection.

However, the reality is that LCMA can still affect devices with such microphone structure alterations. The results of these experiments have been successfully replicated

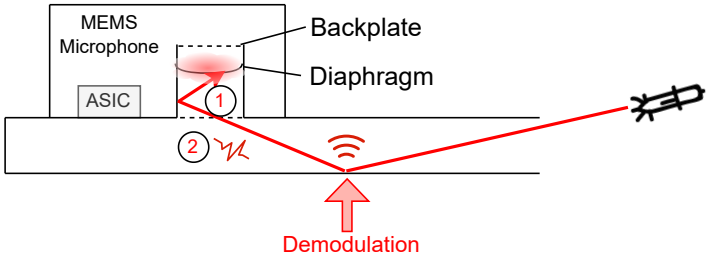


Figure 12: L-Shaped Structure Microphone Attack Scenario

and acknowledged by Vendor-A. We hypothesize that the activated signal consists of two parts: (1)The light energy received by the microphone diaphragm. (2)The modulated laser being demodulated by the wall of sound path, creating an audible command signal that is then collected by the microphone. In section 7.1, we will discuss the root causes of the phenomenon.

Invisibility of LCMA

While laser injection attacks are typically inaudible, their visibility can compromise covert operations. To address this, we experiment with an infrared laser beam, achieving successful attacks while remaining visually undetected. The use of a handheld infrared observation device is crucial for the precise alignment of the laser transmitter array, as shown in Figure 13.

(a) Infrared Observation Device (b) Infrared Laser Injection Attack

Figure 13: Invisible LCMA

The successful implementation of infrared lasers marks a significant step toward truly covert operations by eliminating visual detection risks. However, this method requires specialized equipment, posing a challenge for practical, user-friendly applications in real-world scenarios. Future exploration aims to integrate this technology seamlessly for practical use.

6 Mitigation

This work was first presented at two hacking competitions[2] and successfully attacking provided AI speakers. The organizers of these competitions also informed the affected vendors. The main purpose of this work is to provide defense against potential attacks.

[2]GeekPwn & Tianfu Cup: These competitions provide a platform for security researchers, hackers, and cybersecurity enthusiasts to showcase their skills and discover vulnerabilities in various technologies and systems. Both are highly regarded hacking competitions in China.

We rigorously test LCMA and find it capable of compromising a broad spectrum of smart devices, extending beyond just VC systems. In response, we collaborate with Vendor A to enhance the security of their VC products.

Based on photoacoustic principles eq.3 and the testing results [XZJX21], we propose this new hardware defense strategy, as shown in Figure 14, using three combined mitigation measures: L-shaped structure, light-absorbing material, and optical filter to achieve the defense mission.

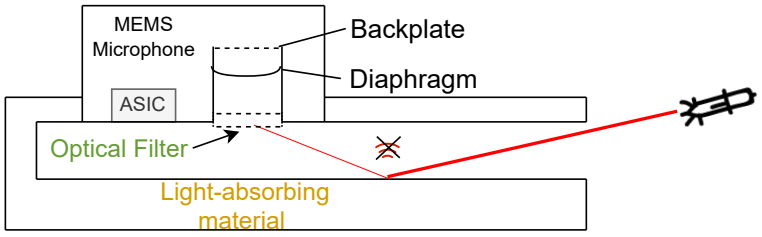


Figure 14: Our Mitigation Strategy

- **L-shaped structure** Given the majority of laser injection attack scenarios (as referenced and discussed in this paper), the optimal attack effectiveness occurs when the laser is directly aimed at the microphone of VC device. Direct aiming allows the microphone diaphragm to receive the maximum energy from the light. In consideration of the physical characteristics of straight-line light propagation and the engineering necessity to maintain microphone sensitivity, we recommend the use of an L-shaped acoustic channel structure to effectively block the energy from directly incident light.
- **Light-absorbing material** To mitigate the potential energy resulting from laser light reflections, caused by laser light reflecting off the elongated walls of the acoustic channel and reaching the internal microphone diaphragm, as well as to address demodulation effects, we opt for a dedicated light-absorbing material to shape the walls of the acoustic channel. The selection of this material should prioritize the minimization of its beta parameter, as well as the parameters β and ν_a, aiming to minimize the generation of photoacoustic signals.
- **Optical filter** Given our experimental findings on light colors, as shown in Table 6, we recommend the integration of a color filter on the exterior surface of the microphone diaphragm. The signal intensity generated by laser irradiation on the MEMs diaphragm is independent of the laser wavelength [SCR+20]. Therefore, we take the example of 450nm laser to illustrate the selection principle of the color filter. The color of the filter should be determined based on the principles of complementary color theory. Specifically, colors with an RGB component value of 0 for the blue (B) component, or simply black (where all RGB components are set to 0), can be chosen to effectively mitigate the impact of all offensive laser emissions.

7 Discussion

This section analyzes the physical roots of LCMA. We propose a notable interpretation of the root cause of laser attacks on MEMs microphones, grounded in physical knowledge and experimental results. We then analyze the reasons why LCMA is so effective and explain why LCMA choose laser arrays over other solutions to address multi-microphone triggered attack scenarios and how LCMA can counter Voice-print Detection. Additionally, we introduce LCMA's limitations.

7.1 Physical Root Causes Analysis

MEMs, ECMs, and piezoelectric microphones are susceptible to vulnerabilities outside their standard human voice frequency range of 35Hz to 1700Hz due to their material composition [SG10], leading to self-demodulation phenomena [HWC+23]. Our experiments on MEMs microphones, using various laser waveforms like square and sine waves, reveal that these microphones only react to changes in laser power. Lasers impact MEMs through mechanical, thermal, and electrical effects, with our research suggesting that the internal photoacoustic effect is the predominant cause of microphone response, overshadowing thermal, mechanical, or photoelectric factors.

Our findings reveal that MEMs microphones react specifically to variations in laser power intensity, confirming that it's the changes in laser light intensity that trigger a responses in these microphones. While the underlying physics of laser attacks align with the established principle of the photoacoustic effect, our work offers a novel interpretation, grounded in both theoretical formulations and empirical results, that advances our understanding of how laser interactions are converted into electrical signals.

Thermal Effect

In our experiment to assess thermal influences, we expose a MEMS microphone to a 450°C soldering iron, simulating rapid and periodic heating within a 1-30mm range. No response is detected, indicating thermal effects don't trigger the microphone. In contrast, we employ a variable-power laser, ensuring a power change frequency of 5Hz and an average power of 700mW, which is significantly lower than the previous temperature. This confirms that the microphone's reaction is not due to thermal effects.

Mechanical Effect

To evaluate mechanical impacts, we consider the ADMP401 microphone's equivalent input noise (EIN) of 32 dBSPL (decibels Sound Pressure Level), the minimum sound intensity it responds to [ADM]. We use a $P_{laser} = 100$mW laser with a 12mm aperture, calculating the light pressure with $P = (1 + R)\frac{P_{laser}}{cS}$, where $R \leq 1$, R is the reflectivity of the microphone's material and S is the laser's aperture area. The resulting sound intensity, determined by $I = 20 \log_{10} \frac{P}{2*10^{-5}}$, is significantly below the EIN threshold at -10.611 dBSPL. Despite this, the microphone still produces an output, indicating that the mechanical effect of laser pressure is not the cause of its response to laser signals.

Electronical Effect

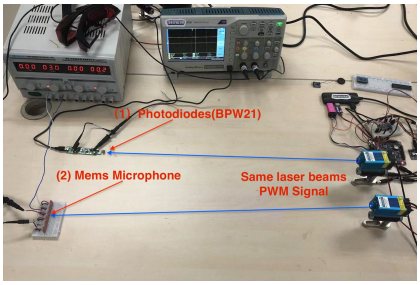


Figure 15: Photodiode VS. MEMs Microphone

To investigate the potential influence of the photoelectric effect, we design a comparative experiment illustrated in Figure 15. In this setup, we utilize the same post amplifier to amplify signals received by both a MEMs microphone module and a BPW21 photodiode, known for its reliance on the photoelectric effect. The experimental arrangements involve exposing both the BPW21 photodiode and the ADMP401 MEMs module to an identical laser signal. Observations made using an oscilloscope connected to these devices show distinctly different responses on the two channels. The disparity in the signals between the photodiode and the MEMs module effectively negates the photoelectric effect as a

plausible explanation for the observed phenomena.

Photoacoustic Effect

The equation for the photoacoustic effect can be expressed as:

$$P(x,t) = \beta \nu_a(x) H(x) \frac{\partial I(x,t)}{\partial t} \tag{3}$$

where:

- P(x,t) represents the photoacoustic signal generated by the material
- β is the isobaric thermal expansion coefficient of the material
- $\nu_a(x)$ is the absorption coefficient at the laser wavelength
- H(x) is the local fluence of the laser beam
- $\frac{\partial I(x,t)}{\partial t}$ denotes the laser beam intensity's temporal profile

Table 7: Photoacoustic Coefficients for Different Materials[BM22]

Material	Photoacoustic Coefficient (W/cm^2*K)
Air	3*10^{-4}
metal	$\leq 1 * 10^{-3}$
water	About 1.5-2.0*10^{-2}
Silicon	0.18
Polystyrene (PS)	About 1-3
PMMA	About 2-4

Our study extends and diverges from prior works like LightCommands by demonstrating that various materials, not just microphone diaphragms, can generate photoacoustic signals [CSF21, Cyr23]. Table 7 illustrates that semiconductor materials and plastics, often used in MEMS and ECM microphone diaphragms, have higher photoacoustic coefficients, making them more susceptible to laser stimulation. Further, our experiments with piezoelectric microphones show that laser stimulation on the sensor's metal backplate produced less sound compared to when the laser is offset to partially hit the sensor's PS plastic case. This suggests that the defense strategy proposed in LightCommands, which involves a movable shading element in front of the microphone diaphragm, may not be entirely effective. The reason being that the movable shading element, when exposed to laser light, could generate photoacoustic signals that are picked up by the MEMs diaphragm, ultimately triggering the VC device.

Additionally, in Section 5.3, we explore the L-shaped structure sound path attack scenario, where the sound path within mobile phones primarily involves the reflection of laser-induced light, triggering the MEMs microphones. Our findings are contrasted with Benjamin Cyr's Ph.D. thesis[Cyr23], providing a broader understanding of laser injection's effectiveness on various capacitive sensors.

7.2 Counter Voice-print Detection

The efficacy of LCMA in IoT environments with voice print detection is underscored by its ability to integrate with advanced audio manipulation techniques. Studies have highlighted vulnerabilities in voice print detection systems, indicating their susceptibility to well-crafted audio inputs, which can compromise their security [YLY23]. LCMA capitalizes on this vulnerability through its laser command injection capabilities. Additionally, the system's effectiveness is further amplified when used in conjunction with replicated or pre-recorded voice samples of the device owner. Techniques for replicating or recording voice samples have been demonstrated to bypass voice authentication protocols effectively [Juz19]. By employing these voice samples, LCMA adeptly circumvents voice print security measures, presenting a significant challenge to the current security paradigm in IoT devices.

7.3 Analysis the Effectiveness of LCMA

Experiments demonstrate that the LCMA can effectively target a wide range of smart devices.

Reasons for transmitter array Our approach employs a transmitter array utilizing Pulse Width Modulation (PWM) and a multi-channel control algorithm. This design offers advantages over methods like enlarging the laser aperture, particularly in scenarios requiring the activation of devices with multiple microphones. Modulated audio signals, due to their distinct spectral features, might be used to differentiate genuine voice from laser-induced signals. However, many Voice-controlled (VC) systems process sound using Digital Signal Processing (DSP) chips, with high-frequency modulation typically occurring on the device side. Due to the spectral characteristics of PWM control signals, the information processed on the device side is insufficient, leading to the inability of cloud-based voice recognition functions to correctly differentiate between genuine audio and light-induced signals. As a result, the issues of sensor fusion can only be solved by LCMA.

Robustness of LCMA test commands LCMA's robustness is evident when handling test voice commands. These commands, typically derived from recorded audio files, are converted into control signals for the laser transmitter using the PWM signal algorithm. This process involves simulating a sine wave through an inertial link's impulse equivalence, resulting in a PWM wave with minimal harmonic components. Thus, LCMA exhibits remarkable resilience against audio signal noise, maintaining effectiveness even amidst environmental interference mixed with voice command audio.

7.4 Limitations

In our study, we acknowledge several limitations that need to be addressed for LCMA to adapt to complex, long-range injection scenarios. Notably, the precision required for aiming, particularly when deploying the system on drones, demands high stability in drone flight to maintain target lock—a challenge that necessitates advanced automation in targeting capabilities. Other constraints include the development of a more user-friendly standalone integrated testing suite that would enable simpler operations, such as one-touch recording and command injection, without the need for a connected computer to supply the injection commands. Future work should focus on these aspects to enhance the usability and effectiveness of LCMA in various operational contexts.

8 Conclusions

In this paper, we propose LCMA, a new laser-based audio injection attack approach for Voice-controlled (VC) systems. Our approach utilizes Pulse Width Modulation (PWM) as a voice-to-laser conversion modulation method, where the lasers are replicated onto a multi-channel laser rack to form a laser array. This strategy effectively addresses complex Sensor Fusion challenges while LightCommands can not solve this problem. Moreover, our solution eliminates the need for additional signal controllers for the lasers, allowing LCMA to be easily extended for controlling multiple lasers in attacks on microphone arrays. Through experiments on various types of VC devices, we demonstrate that LCMA can successfully take over the new VC devices with microphone arrays and subsequently control the concomitant IoT devices.

Acknowledgement

This work was supported in part by National Key Technologies R & D Program of China under Grant No.20221880004 and NSFC under Grant 62132011. We thank Prof. Jackie Mao and anonymous reviewers for their comments to improve the paper. Jianwei Zhuge is the corresponding author.

References

[Aar03] Parham Aarabi. The fusion of distributed microphone arrays for sound localization. *EURASIP Journal on Advances in Signal Processing*, 2003:1–10, 2003.

[ADM] In *https://www.analog.com/media/en/technical-documentation/obsolete-data-sheets/ADMP401.pdf*.

[AWS+19] Eirini Anthi, Lowri Williams, Małgorzata Słowińska, George Theodorakopoulos, and Pete Burnap. A supervised intrusion detection system for smart home iot devices. *IEEE Internet of Things Journal*, 6(5):9042–9053, 2019.

[BJC15] Jakub Breier, Dirmanto Jap, and Chien-Ning Chen. Laser profiling for the back-side fault attacks: with a practical laser skip instruction attack on aes. In *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*, pages 99–103, 2015.

[BM22] Rita Clarisse Silva Barbosa and Paulo M Mendes. A comprehensive review on photoacoustic-based devices for biomedical applications. *Sensors*, 22(23):9541, 2022.

[BW01] Michael Brandstein and Darren Ward. *Microphone arrays: signal processing techniques and applications*. Springer Science & Business Media, 2001.

[CBR+13] Pritesh V. Chhajed, Mugdha A. Bondre, Vaibhav M. Rekhate, Pushkar C. Chaudhari, Priyanka G. Aher, and S.P. Metkar. Humanizing the interface: Voice activated devices. In *2013 Texas Instruments India Educators' Conference*, pages 243–247, 2013.

[CSF21] Benjamin Cyr, Takeshi Sugawara, and Kevin Fu. Why lasers inject perceived sound into mems microphones: Indications and contraindications of photoacoustic and photoelectric effects. In *2021 IEEE Sensors*, pages 1–4. IEEE, 2021.

[Cyr23] Benjamin Cyr. *Characterizing Laser Signal Injection and its Impact on the Security of Cyber-Physical Systems*. PhD thesis, 2023.

[DAY22] Donghui Dai, Zhenlin An, and Lei Yang. Inducing wireless chargers to voice out for inaudible command attacks. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 503–520. IEEE Computer Society, 2022.

[DLZZ14a] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pages 63–74, 2014.

[DLZZ14b] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, SPSM '14, page 63–74, New York, NY, USA, 2014. Association for Computing Machinery.

[dVIV+17] Lara del Val, Alberto Izquierdo, Juan José Villacorta, Luis Suárez, et al. Using a planar array of mems microphones to obtain acoustic images of a fan matrix. *Journal of Sensors*, 2017, 2017.

[Gol92] Jason M Goldberg. *Signal processing for high resolution pulse width modulation based digital-to-analogue conversion*. PhD thesis, University of London, 1992.

[hara] Stm32f103cb. https://www.st.com/en/microcontrollers-microprocessors/stm32f103cb.html. Accessed: 2024-01-15.

[harb] Stm32f407ve. https://www.st.com/en/microcontrollers-microprocessors/stm32f407ve.html. Accessed: 2024-01-15.

[harc] Stm32h7 pwm. https://teamgloomy.github.io/stm32h7_pwm.html. Accessed: 2024-01-15.

[HWC+23] Peng Huang, Yao Wei, Peng Cheng, Zhongjie Ba, Li Lu, Feng Lin, Fan Zhang, and Kui Ren. Infomasker: Preventing eavesdropping using phoneme-based noise. In *NDSS*, 2023.

[Juz19] R Juzenaite. "security vulnerabilities of voice recognition technologies, 2019.

[KMR12] Kenichi Kumatani, John McDonough, and Bhiksha Raj. Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors. *IEEE Signal Processing Magazine*, 29(6):127–140, 2012.

[MBM+18] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, and Yuval Elovici. N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, 17(3):12–22, 2018.

[MZL20] Jian Mao, Shishi Zhu, and Jianwei Liu. An inaudible voice attack to context-based device authentication in smart iot systems. *Journal of Systems Architecture*, 104:101696, 2020.

[Nyq] Nyquist frequency. https://en.wikipedia.org/wiki/Nyquist_frequency. Accessed: 2024-01-15.

[OAV04] Alejandro R Oliva, Simon S Ang, and Thuy V Vo. A multi-loop voltage feedback filterless class-d switching audio amplifier using unipolar pulse-width-modulation. *IEEE transactions on consumer electronics*, 50(1):312–319, 2004.

[RHRC17] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Making microphones hear inaudible sounds. MobiSys '17, page 2–14, New York, NY, USA, 2017. Association for Computing Machinery.

[San93] MB Sandler. Digital-to-analogue conversion using pulse width modulation. *Electronics & Communication Engineering Journal*, 5(6):339–348, 1993.

[SCCM20] Jiachen Sun, Yulong Cao, Qi Alfred Chen, and Z. Morley Mao. Towards robust LiDAR-based perception in autonomous driving: General black-box adversarial sensor attack and countermeasures. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 877–894. USENIX Association, August 2020.

[SCR+20] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. Light commands: Laser-Based audio injection attacks on Voice-Controllable systems. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 2631–2648. USENIX Association, August 2020.

[SG10] Jan G Svec and Svante Granqvist. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, 19(4):356–368, 2010.

[SKKK17] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 445–467. Springer, 2017.

[SM17] Liwei Song and Prateek Mittal. Poster: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 2583–2585, 2017.

[TLG+15] Shahin Tajik, Heiko Lohrke, Fatemeh Ganji, Jean-Pierre Seifert, and Christian Boit. Laser fault attack on physically unclonable functions. In *2015 Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC)*, pages 85–96, 2015.

[VZSS15] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: Exploiting the gap between human and machine speech recognition. WOOT'15, page 16, USA, 2015. USENIX Association.

[WM21] Ganyu Wang and Miguel Vargas Martin. Segmentperturb: Effective black-box hidden voice attack on commercial asr systems via selective deletion. In *2021 18th International Conference on Privacy, Security and Trust (PST)*, pages 1–12, 2021.

[XZJX21] Zhijian Xu, Guoming Zhang, Xiaoyu Ji, and Wenyuan Xu. Evaluation and defense of light commands attacks against voice controllable systems in smart cars. *Noise & Vibration Worldwide*, 52(4-5):113–123, 2021.

[YLY23] Baochen Yan, Jiahe Lan, and Zheng Yan. Backdoor attacks against voice recognition systems: A survey. *arXiv preprint arXiv:2307.13643*, 2023.

[YLZ+20] Qiben Yan, Kehai Liu, Qin Zhou, Hanqing Guo, and Ning Zhang. Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided waves. In *Network and Distributed Systems Security (NDSS) Symposium*, 2020.

[YXL16] Chen Yan, Wenyuan Xu, and Jianhao Liu. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle. *Def Con*, 24(8):109, 2016.

[ZYJ+17] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 103–117, New York, NY, USA, 2017. Association for Computing Machinery.