Poster Abstract: LidarPhone: Acoustic Eavesdropping using a Lidar Sensor

Sriram Sami  
National University of Singapore  
srirams@comp.nus.edu.sg

Sean Rui Xiang Tan  
National University of Singapore  
seantanr@comp.nus.edu.sg

Yimin Dai  
National University of Singapore  
e0505408@u.nus.edu

Nirupam Roy  
University of Maryland, College Park  
nirupam@cs.umd.edu

ABSTRACT

Private conversations are an attractive target for malicious actors intending to conduct audio eavesdropping attacks. Previous works discovered unexpected vectors for these attacks, such as analyzing high-speed video of objects adjacent to sound sources, or using Wi-Fi signal information. We propose LidarPhone, a novel side-channel attack that exploits the lidar sensors in commodity robot vacuum cleaners to perform acoustic eavesdropping attacks. LidarPhone is able to detect the minute vibrations induced on objects that are near audio sources, and extract meaningful signals from inherently noisy raw lidar returns. We evaluate a realistic scenario for potential victims: recovering privacy-sensitive digits (e.g., credit card numbers, social security numbers) emitted by computer speakers during teleconferencing calls. We implement LidarPhone on a Xiaomi Roborock vacuum cleaning robot and perform a comprehensive series of real-world experiments to determine its performance. LidarPhone achieves up to 91% accuracy for digit classification.

CCS CONCEPTS

- Computer systems organization → Sensors and actuators;
- Security and privacy → Embedded systems security.

KEYWORDS

Lidar, Acoustic Side-Channel, Eavesdropping

ACM Reference Format:


1 INTRODUCTION

Smart sensing devices are increasingly ubiquitous in modern homes, and have provided many opportunities for acoustic-side-channel attacks on private conversations. Devices containing microphones such as smart speakers, baby monitors, and smart security cameras are considered the usual threats to acoustic privacy. However, recent work also demonstrates that other types of sensors (e.g., accelerometers, gyroscopes) may also pose similar threats [5, 8], enabling more devices to be exploited as microphones. In this work, we propose LidarPhone [11], a novel acoustic side channel attack from a seemingly harmless household appliance – a vacuum cleaning robot. LidarPhone repurposes inexpensive Light Detection and Ranging (lidar) sensors, used aboard newer vacuum robots to determine distances to surrounding obstacles for navigation [1], to collect acoustic signals from the environment.

Fundamentally, LidarPhone senses vibrations that are known to be induced on objects close to sound sources [2]. Based on this concept, Figure 1 demonstrates a potential attack scenario. An adversary launches a remote software attack on the vacuum cleaner (recently observed to be possible [4]), and obtains a raw lidar sensor data stream. The stream is transmitted to the cloud for the remote adversary to process and reconstruct the source audio.

LidarPhone’s eavesdropping technique appears similar to that of a laser mic [9]. Laser mics target lasers at highly (i.e., specularly) reflective surfaces such as windows and mirrors, and process the focused reflected beam to eavesdrop on audio near the target surface. While designed for a different purpose, lidars possess both a laser transmitter and receiver, seemingly creating the potential for them to act as a laser mic. However, an attacker can launch the LidarPhone attack remotely, which relies on different physical principles. Specifically, lidars on robot vacuum cleaners are designed to only process signals from diffusely reflecting surfaces (which spread the reflected light equally in all directions) as most household objects do not produce specular reflections.

Figure 1: An example of a LidarPhone attack. The adversary remotely exploits the lidar sensor on a victim’s robot vacuum cleaner to recover privacy sensitive information (e.g., credit card numbers) during a teleconferencing meeting.
These diffuse reflections are relatively unfocused and low in intensity, causing the reflected signals to have an extremely low signal-to-noise ratio (SNR). Therefore, LidarPhone implements signal processing techniques to increase the signal’s effective SNR. These include: noise profiling followed by spectral subtraction to reduce noise, and equalization to emphasize lower frequency components since objects tend to attenuate higher frequencies.

Furthermore, lidars commonly rotate at 300 RPM (5 Hz), and can therefore only sample a given point on a target surface at a 5 Hz sampling rate. We halt the lidar’s rotation to increase the sampling rate by a factor of the number of samples per rotation. With robot vacuum lidars operating at a typical rate of 360 samples per rotation, LidarPhone achieves a 1.8 kHz (5 Hz × 360) sampling rate, which is lower than the 5 kHz required to recover an intelligible speech signal [10]. We evaluate LidarPhone with real-world experiments and achieve 91% accuracy in identifying digits (e.g., “one”, “two”) from LidarPhone-recovered audio.

2 LIDARPHONE ATTACK DESIGN

We present the design of LidarPhone in order to recover (1) sensitive digit information such as credit card numbers; (2) the gender of the speaker; or (3) the speaker’s identity. The attacker’s goal is to successfully conduct a stealthy remote eavesdropping attack using lidar readings from a victim’s robot vacuum cleaner. The attacker has the capabilities to remotely control the robot, stop the lidar from rotating, and obtain raw lidar intensity values. Additionally, the digit inference attack targets a specific victim (e.g., celebrities), and the attacker has the capability to train on their recorded speech.

Figure 2 depicts LidarPhone’s design overview. The attacker first pre-processes the raw signals, removing any DC offset, outliers, and noise. Since objects primarily attenuate high frequencies, the attacker equalizes the signal to amplify low frequency components. The attacker then converts each processed time-series signal into a 200 × 200 spectrogram image that is used as input to a convolutional neural network (CNN) classifier. Spectrograms represent a signal’s frequency and temporal information compactly as an image, and CNNs are architected to achieve high accuracy for image classification tasks [7]. During the bootstrapping phase, the attacker trains the CNN classifier with the pre-processed spectrograms and ground-truth labels. In the attack phase, the attacker tests the captured signals against the trained model to output a final predicted digit, gender, or speaker identity.

3 FEASIBILITY STUDY

We implement LidarPhone on a popular robot vacuum cleaner, the Xiaomi Roborock S5, and measure its effectiveness on multiple tasks. We play audio of spoken digits (e.g., “one”, “two”) from a Logitech Z623 speaker-subwoofer system, near a common household object (trashcan), which is targeted by the lidar. We vary both environmental and system parameters to obtain more than 19 hours (30K utterances) of recorded audio. We use the Free Spoken Digit Dataset as the source audio [6] for digit classification, and TIDIGITS [3] for gender and speaker classification. LidarPhone achieves 91%, 96%, and 67% accuracy for the digit, gender, and speaker classification tasks respectively.

Figure 3: Figure depicts LidarPhone’s overall performance for each evaluated task.

ACKNOWLEDGMENTS

This research was partially supported by a grant from Singapore Ministry of Education Academic Research Fund Tier 1 (R-252-000-A26-133).

REFERENCES