Lightbox: Sensor Attack Detection for Photoelectric Sensors via Spectrum Fingerprinting

DOHYUN KIM, MANGI CHO, HOCHEOL SHIN, JAEHOON KIM, JUHWAN NOH, and YONGDAE KIM, KAIST, Republic of Korea

Photoelectric sensors are utilized in a range of safety-critical applications, such as medical devices and autonomous vehicles. However, the public exposure of the input channel of a photoelectric sensor makes it vulnerable to malicious inputs. Several studies have suggested possible attacks on photoelectric sensors by injecting malicious signals. While a few defense techniques have been proposed against such attacks, they could be either bypassed or used for limited purposes.

In this study, we propose Lightbox, a novel defense system to detect sensor attacks on photoelectric sensors based on signal fingerprinting. Lightbox uses the spectrum of the received light as a feature to distinguish the attacker's malicious signals from the authentic signal, which is a signal from the sensor's light source. We evaluated Lightbox against (1) a saturation attacker, (2) a simple spoofing attacker, and (3) a sophisticated attacker who is aware of Lightbox and can combine multiple light sources to mimic the authentic light source. Lightbox achieved the overall accuracy over 99% for the saturation attacker and simple spoofing attacker, and robustness against a sophisticated attacker. We also evaluated Lightbox considering various environments such as transmission medium, background noise, and input waveform. Finally, we demonstrate the practicality of Lightbox with experiments using a single-board computer after further reducing the training time.

CCS Concepts: • **Security and privacy → Embedded systems security**; • **Hardware → Sensors and actuators;**

Additional Key Words and Phrases: Photoelectric sensors, neural networks, signal fingerprinting

ACM Reference format:
Dohyun Kim, Mangi Cho, Hocheol Shin, Jaehoon Kim, Juhwan Noh, and Yongdae Kim. 2023. Lightbox: Sensor Attack Detection for Photoelectric Sensors via Spectrum Fingerprinting. *ACM Trans. Priv. Sec.* 26, 4, Article 46 (October 2023), 30 pages.
https://doi.org/10.1145/3615867

1 INTRODUCTION

Sensors are widely used in cyber-physical systems such as autonomous vehicles, medical devices, and industrial control systems. In such systems, sensors transform information from the physical

D. Kim, M. Cho, H. Shin, and Y. Kim contributed equally to this research.

The authors gratefully acknowledge the support from Nano UAV Intelligence Systems Research Laboratory at Kwangwoon University, originally funded by Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD).

Authors' addresses: D. Kim, M. Cho, H. Shin, J. Kim, J. Noh, and Y. Kim, KAIST, Republic of Korea; e-mails: dohyunjk@kaist.ac.kr, mgcho0608@kaist.ac.kr, h.c.shin@kaist.ac.kr, jaehoon.kim99@kaist.ac.kr, juwhan@kaist.ac.kr, yongdaek@kaist.ac.kr.

2471-2566/2023/10-ART46 $15.00
https://doi.org/10.1145/3615867

world to the digital domain. Sensors are typically categorized into two: passive and active sensors. Passive sensors collect information about a measurand solely based on physical signals originating from it, while active sensors operate by emitting a physical signal and measuring the response from a measurand. Radars and lidars are common examples of active sensors used to estimate directions and distances to objects without physical contact.

Photoelectric sensors are a common type of active sensors. They are typically composed of a light transmitter and a photoelectric receiver and are used to measure the distance to an object or the presence of an object using light. They are extensively used for applications in various areas, such as surveillance [1, 2], medical devices [16, 47], autonomous driving [49], industrial safety [15, 38], and theme park safety [28]. In these applications, they are used for important tasks such as object detection [7, 9], monitoring production and packaging lines in factories [6, 8], detecting the presence of human bodies in hazardous areas [15], detecting the leakage of hazardous liquids from pipes inside valve boxes [5], and the measurement of product filling rates within transparent containers [10].

Despite their versatility, photoelectric sensors can be vulnerable to security threats due to the public exposure of their optical receivers. Since photoelectric sensors operate by detecting incoming light signals, an attacker can manipulate the sensor's measurements by injecting false signals from other light sources similar to the authentic source. This allows adversaries to interfere with authentic optical signals using (1) light sources of other models capable of generating signals of a similar wavelength or even (2) light sources of the same model as the authentic one.

Several studies have highlighted the security risks faced by photoelectric sensors. For instance, Shin et al. demonstrated that attackers could trigger false fire alarms or suppress real ones by saturating and spoofing optical beam smoke detectors (OBSDs) [34]. Similarly, Park et al. showed that drop sensors in medical infusion pumps could be spoofed by injecting false light signals [27]. While optical shielding around vulnerable sensors can prevent such exploits by blocking all external light, this solution is not feasible for all applications. Some applications, such as OBSDs, cannot be optically shielded, because the shielding could obstruct the detection of smoke, the intended measurand. Similarly, implementing shielding in applications designed to detect external changes, such as outdoor beam detectors [1] or industrial hazard monitoring systems [15], is also impractical. PyCRA, proposed by Shoukry et al., aimed to detect sensor spoofing attacks on general active sensors by randomly introducing sudden notches into the emitted signal [36]. Unfortunately, Shin et al. demonstrated that PyCRA could be bypassed unless the sensor employs impractically high sampling rates [35].

In this study, we propose *Lightbox*, a novel photoelectric sensor attack detection system. Lightbox utilizes the color consistency problem [22, 32], where products of the same light source models have different spectrum distributions, even when they are manufactured in the same production line. It differentiates the sensor's light signal from other incoming signals by fingerprinting the spectrum of the light source using a machine-learning technique. Its training process is composed of the following steps: (1) collect the light spectrum of authentic and reference light sources, which are the additional light sources of the same model as the authentic light source, (2) train a feature extractor that can classify light sources of reference datasets, (3) train a one-class (OC) classifier using a pre-trained feature extractor. Here, the OC classifier is used to classify unseen light signals. After these training steps, Lightbox can distinguish incoming light signals from authentic light signals.

To evaluate the performance of Lightbox, we conducted experiments using LEDs and lasers of various wavelengths. Before proceeding with the evaluation, we defined common metrics including accuracy, precision, and recall. We also presented the minimum precision and recall to comprehensively test whether Lightbox has any bias. Next, we evaluated Lightbox's detection

performance against various types of sensor attacks. These included (1) a saturation attack, (2) a simple spoofing attack, and (3) a sophisticated attack that utilizes multiple light sources to mimic the authentic one. Our experimental results demonstrated that saturation attacks and simple spoofing attacks (using a light source of the same model as the authentic source) could be detected by Lightbox with an accuracy rate exceeding 99%. Furthermore, we found that Lightbox remained effective, with a detection accuracy of over 99%, even against a sophisticated attacker using a small number of light sources. We also examined whether Lightbox's detection performance remained stable amid practical variables such as the background light noise, transmission medium, and so on. Last, we assessed the long-term robustness of Lightbox. The details of our evaluation settings and results are further elaborated in Section 5.

After evaluations, we analyzed the interpretability of Lightbox to find out which features were important in the classification of Lightbox. Also, to demonstrate the practicality of Lightbox, we suggested a fine-tuning method to reduce the training efforts of Lightbox. Moreover, we implemented Lightbox on Jetson TX2 to check out the training and operating duration in real-world systems. Our contributions are summarized as follows:

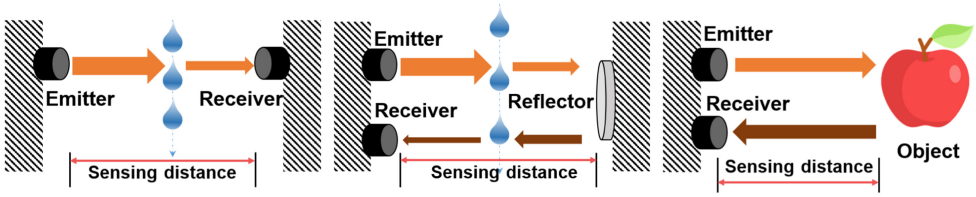
- We propose Lightbox, a fingerprinting-based photoelectric sensor attack detection system, which protects sensors from adversarial signal injection attacks by analyzing the received light spectrum. To the best of our knowledge, this is the first proposal to apply the principle of fingerprinting on light sources to detect sensor attacks.
- We evaluated the detection performance of Lightbox via OC classification using seven different types of light sources. First, we verified the capability of Lightbox to differentiate the authentic light source from the other sources of the same model and detect saturation and spoofing signals. We then evaluated the influence of several factors that were likely to affect the performance of Lightbox, considering longer-term consistency, background noise, input current, transmission medium, and relative position.
- We analyzed the interpretability of the classification results obtained by Lightbox. The application of the Local Interpretable Model-Agnostic Explanations (LIME) structure and feature selection revealed the importance of the spectrum around the peak.
- We verified the operation of Lightbox in a real-world application environment using a single-board computer. We verified that the inefficiency in the training process of Lightbox could be significantly reduced with negligible decreases in the detection accuracy using a fine-tuning method [29]. Finally, through a demo video, we present an end-to-end application scenario of Lightbox, showing its feasibility for deployment.

The remainder of this article is organized as follows: Background on sensor attacks and the color consistency problem is described in Section 2. In Section 3, we introduce the threat model, design goal, and system overview of Lightbox. The implementation of Lightbox is presented in Section 4, followed by the evaluation results and interpretability discussions in Sections 5 and 6. A discussion on optical signal-to-noise ratio and accuracy with respect to the distance, practicality, end-to-end application scenario, and limitations of Lightbox is presented in Section 7. The related works and conclusions are provided in Sections 8 and 9.

2 BACKGROUND

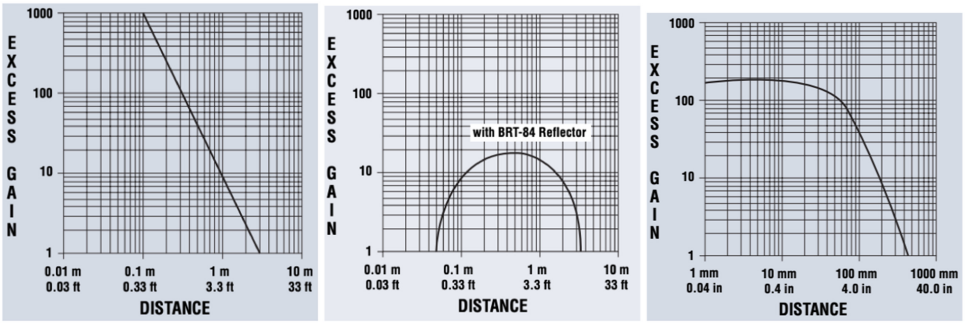
2.1 Photoelectric Sensors

Sensors can be broadly classified into two categories, namely, passive and active. Passive sensors collect information about a measurand solely based on physical signals originating from it. Active sensors operate by emitting a physical signal and measuring the response from a measurand.



(a) Through-beam (b) Retro-reflective (c) Diffuse-reflective

Fig. 1. Three types of photoelectric sensors.



(a) Through-beam (b) Retro-reflective (c) Diffuse-reflective

Fig. 2. Excess gain as a function of distance.

Photoelectric sensors, one of the popular active sensors, are used to measure the distance to an object or the presence of an object using light.

Photoelectric sensors consist of a light source and a photoreceiver. The fundamental operating principle of photoelectric sensors involves calculating the difference between the amount of light transmitted by the emitter and the amount of light captured by the receiver. Nowadays, light sources employ light emitting diodes (LEDs), which have a solid-state. Due to their solid-state design, LEDs last for years, are resistant to damage, smaller in size compared to incandescent lights, and capable of withstanding a wide range of temperatures.

Photoelectric sensors are divided into three types of sensing modes based on the arrangement of components: through-beam, retro-reflective, and diffuse-reflective [26]. The three sensing modes of photoelectric sensors are depicted in Figure 1. These modes differ in terms of the spatial arrangement of the transmitter and receiver, which results in varying excess gain, a significant property of photoelectric sensors. Excess gain refers to the amount of sensing energy obtained by the receiver in excess of what is necessary for the operation of the sensor. Figure 2 illustrates the correlations between sensing distance and excess gain in the three sensing modes [4]. Note that the sensing distance of a photoelectric sensor can vary, depending on its sensing mode or application type, and the sensing distance refers to the maximum distance at which the photoelectric sensor can perform as intended. In other words, the photoelectric sensor can adjust its usage distance within this sensing distance.

A through-beam sensor consists of an emitter and a receiver that are located on opposite sides, with the object being measured passing between them. The through-beam sensor's sensing distance is the length between the emitter and the receiver, and its excess gain is proportional to the inverse square of the sensing distance. Since the light from the emitter moves directly to the receiver without reflecting, the through-beam sensor is the most efficient sensing mode. Therefore,

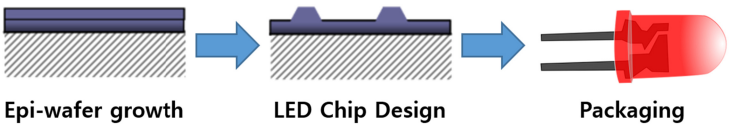


Fig. 3. LED manufacturing process.

through-beam sensors can be employed in applications requiring high excess gain, such as long-range sensing, sensing in unclean environments, and small object detection due to their high efficiency.

In a retro-reflective sensor, the emitter and receiver are located on the same side, with an additional retro-reflector on the opposite side. The sensing distance of a retro-reflective sensor is the distance between the emitter and the reflector. Its excess gain grows as the distance increases until a certain point, then reduces as the distance increases. Furthermore, the size of a retro-reflector affects the excess gain of a retro-reflective sensor. In other words, a larger retro-reflector provides a greater excess gain curve. Retro-reflective sensors have lower excess gains than through-beam sensors, which limits their use for detecting objects.

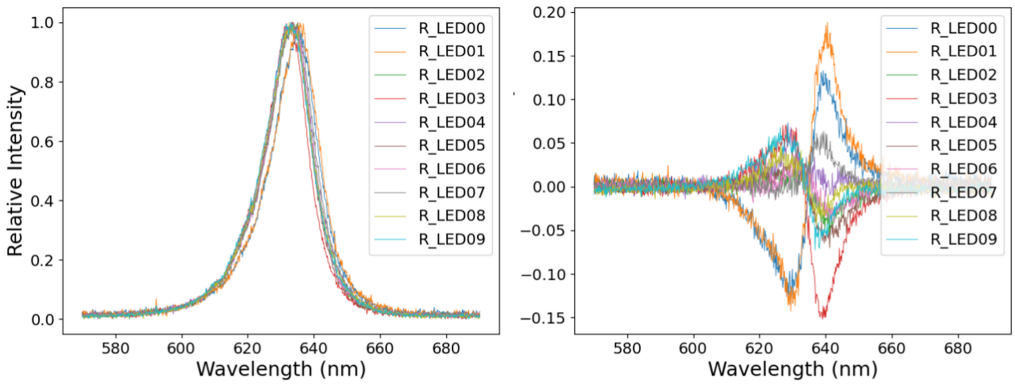
The operation and components of a diffuse-reflective sensor are similar to those of a retro-reflective sensor, because it has an emitter and receiver on the same side, and it emits a signal from the emitter while measuring the fraction of light reflected by a target object. The diffuse-reflective sensor's components, however, do not need to be fixed, although the components of the other mode sensors do. For example, lidars and Time-of-Flight cameras, which are famous diffuse-reflective sensors, identify an object and its distance by measuring the time between when a light signal is sent and when it is detected after being reflected by the object. The sensing distance of a diffuse-reflective sensor is the distance between the emitter and the object. Its excess gain curve is similar to that of a through-beam sensor; however, it has the shortest sensing distance. Since the diffuse-reflective sensor has a low excess gain, its performance is highly dependent on the surrounding environment (e.g., performance degradation of lidars under rain or snow conditions).

2.2 Manufacturing Variability and Color Consistency Problem

Even though two products are manufactured using the same process (e.g., the same machine, mask, and wafer), they are physically distinct [18, 46]. This distinction is caused by uncontrollable and inherent manufacturing variability and randomness, which also applies to light sources. Consequently, attackers or even manufacturers are unable to duplicate a specific product to create fully identical products.

In this section, we explain how difficult it is to control the manufacturing variability of light sources through the manufacturing process of an LED, one of the most common light sources. Figure 3 shows the manufacturing process of an LED. The first step of LED manufacturing process is to fabricate a wafer. The substrate that corresponds to generate the desired color is chosen by the manufacturer and used to build the semiconductor wafer. Epitaxial growth is the most popular approach for creating high-quality mono-semiconductor wafers. After the wafer is synthesized, all parts except the predetermined circuit pattern are cut off using an etching procedure. Once the etching is completed, several semiconductor chips are created on the wafer and isolated from each other. Finally, the chip is phosphor-coated and packed to create the final LED product. All of these procedures, (1) epitaxial growth, (2) chip designing operations (e.g., etching), and (3) phosphor coating, may involve manufacturing variability.

Epitaxial growth aims to grow a single semiconductor material in the same direction as the crystal axis of the underlying single substrate to create the desired layer. This approach reduces defects in the wafer such as holes, lattice mismatches, and dislocations. However, It can introduce significant defects into the epitaxial layer [17]. For instance, poor matching of lattice



(a) Normalized spectrum (b) Difference of the normalized spectrum

Fig. 4. Normalized spectrum and spectrum difference between 10 LEDs within a same red LED model.

characteristics and thermal expansion coefficients result in a high density of threading dislocations in the order of 10^{10}cm^{-2} [48]. It can be reduced to a certain degree; however, it cannot be completely removed [39, 50].

The surface of an etched wafer should ideally be uniformly flat during the chip design process; unfortunately, the surface is uneven and contains faults with a density of $7.5 \cdot 10^{8}\text{cm}^{-2}$ [39]. Furthermore, specific patterned textures are inserted onto the wafers' surfaces to improve luminous efficiency. However, this results in variances in the physical features of the devices, because no two light sources can have the same pattern texture [53].

The volume of the encapsulating phosphor varies for each device due to intrinsic inconsistencies in the phosphor dispensation process and ambient variations, which affect the LED optical efficiency and quality during the phosphor-coating process [37]. Moreover, the light efficiency is influenced by the lens curvature and phosphor concentration inconsistencies [45].

To conclude this section, several factors contribute to the manufacturing variability in LEDs. Other light sources, such as lasers, require complicated manufacturing processes, resulting in uncontrollable manufacturing variability. These manufacturing variabilities alter the electronic characteristics of the device [17], resulting in spectrum discrepancies between light sources within a given model, commonly known as a color consistency problem [22, 32]. Indeed, when analyzing the spectrum of 10 red LED within a same light source model, it can be observed that there exist a difference in the spectrum as shown in Figure 4. Note that the data in Figure 4(b) is measured by the difference between the normalized spectrum of single LED product and the average of the normalized spectrum of all LED products.

3 THREAT MODEL

In this section, we introduce the types of sensor attacks on photoelectric sensors and attacker assumptions.

3.1 Sensor Attacks

There are two types of attacks against photoelectric sensors: saturation and spoofing. Saturation attacks exploit the transition curve of a sensor, which characterizes its input-output relationships (Figure 5). The output has a linear relationship with the input within the desired input range, i.e., the linear region. However, if the input surpasses the linearity limit, then the output cannot keep up with the input, resulting in a flattened curve, which indicates saturation. Attackers can exploit this phenomenon to make sensors virtually oblivious to changes in the input. When the sensor is

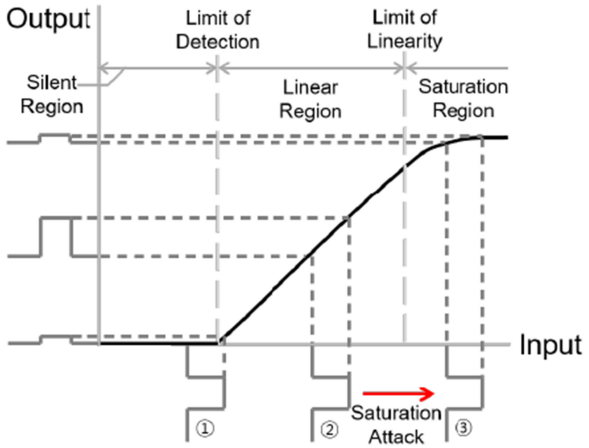


Fig. 5. Typical transition curve of the sensor input and output. The input range is divided into *silent*, *linear*, and *saturation* regions [35]. The objective of a saturation attack is to move the sensor's state to a saturation region (red arrow).

exposed to a significant input offset, the majority of the input variation is suppressed in the output (②→③ in Figure 5).

Spoofing attacks cause fake outputs in the sensor by imitating authentic input signals. For example, attackers can inject arbitrary objects into the radar output by exposing the sensor to fake echoes that resemble authentic echoes at the proper times [12]. This exploits the gap between reality and the shape of received signals. In the case of a photoelectric sensor, an attacker can spoof the intensity of the received light by injecting light using an additional light source that the target sensor can receive.

3.2 Attacker Assumption

We assume the following conditions for the attacker:

No direct access to the target. An adversary cannot directly access the target device. Therefore, the adversary cannot physically damage the victim's device or eavesdrop on the receiving information by tampering with the hardware and software of the target device.

Pre-knowledge of the target. An adversary becomes aware of the target device by using another device of the same model. For example, the adversary knows the model name, manufacturer, and manufacturing steps of the light source used in the target device.

In this work, we focus on an attacker who attempts to manipulate the output of the target sensor via sensor saturation or spoofing attacks. Note that we do not consider a side-channel attacker [21, 40] who uses a non-target stimulus to attack the photoelectric sensors. For the sensor saturation and spoofing attack, the goal of the attacker is to inject fake signals by using malicious light sources that generate light signals that can be received by the target sensor. Since the attacker cannot directly touch the target device, the attacker cannot obtain the benign spectrum received by the target sensor. Instead, the attacker can collect the spectrum of the target sensor's light source in other locations. A strong attacker who is aware of Lightbox can either combine existing light sources or use spectrum modulation to reshape the spectrum to generate the desired spectrum. However, using spectrum modulation on the terahertz frequency band is difficult and expensive due to the complex optical setups required [19, 41]. Therefore, we consider a strong attacker who can generate the desired spectrum by combining multiple light sources.

4 LIGHTBOX

This section introduces the Lightbox design and implementation. The training procedure for the Lightbox is described in detail, followed by a description of the data collection steps.

4.1 Design Goals

We focus on a system capable of detecting malicious optical inputs received by photoelectric sensors based on the threat model (Section 3.2). The design goals of the proposed sensor attack detection system include:

- **G1.** The system should be able to detect sensor attacks including saturation attack and spoofing attack. Furthermore, it should be effective against strong attacker who is aware of our system and can generate sophisticated malicious signals using multiple light sources.
- **G2.** To be compatible with most photoelectric sensor applications, the system should use generic properties, instead of application-specific information.
- **G3.** The system should be able to distinguish authentic light sources from malicious ones even under changes in environmental conditions such as noise.
- **G4.** The sensor attack detection accuracy of the system should not degrade over time.

4.2 System Overview

We present Lightbox, a sensor attack detection system that uses machine learning to achieve the design goals. Lightbox detects sensor attacks by distinguishing the target sensor's transmitter from any other light source, including the attacker's. Lightbox adapts the OC classifier model proposed by Perera et al. [29]. This model is a high-performance OC image classifier that employs a feature extractor to discriminate between significantly similar data and identifies authenticity for various types of objects. The reasons for using a feature extractor and SVM are as follows: First, for the purpose of Lightbox, which is to detect light signal injection attacks, a one-class classifier that distinguishes between data used for training and data not used for training was needed. SVM is one of the models that can perform this task most effectively. However, when we tried the SVM alone, we found that the differences in spectra between different light sources within the same model were either too subtle to distinguish or led to overfitting. Therefore, a preprocessing step that can differentiate the subtle differences between spectrum signals was needed, and the feature extractor was able to perform this role. Consequently, authentic light sources may be distinguished from unseen and similar light sources with appropriate training data.

The workflow of Lightbox comprises operating and training procedures as depicted in Figure 6. The training procedure runs only one time before the operation. The following is a summary of the Lightbox training procedure (implementation details are discussed in Section 4.4):

(i) **Extractor Training Procedure** In addition to the authentic light source, light spectrum samples from the reference light sources, which are extra light sources of the same model as the authentic light source, are gathered and utilized to train the feature extractor.
(ii) **OC Training Procedure** Following the extractor training, an OC classifier is trained using the feature extractor network and the spectrum of the authentic light source.

In the operation procedure, the target photoelectric sensor receives light and obtains spectrum data using a spectrometer. Following that, Lightbox uses the pre-trained feature extractor to extract a feature from the spectrum data and uses the pre-trained OC classifier to verify its authenticity. Based on the result of Lightbox, the back-end system determines whether to trust sensor data or not.

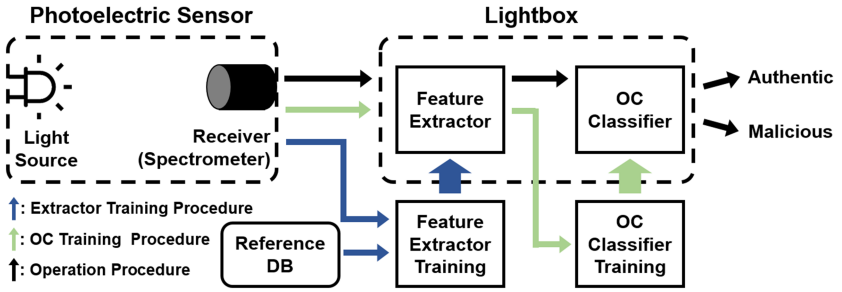


Fig. 6. Workflow of Lightbox training and operation; the operation, extractor training, and OC training procedures are indicated by black, blue, and green arrows, respectively. The thin and thick arrows indicate the flow of each process and the final outputs of the training procedure, respectively.

4.3 Data Collection

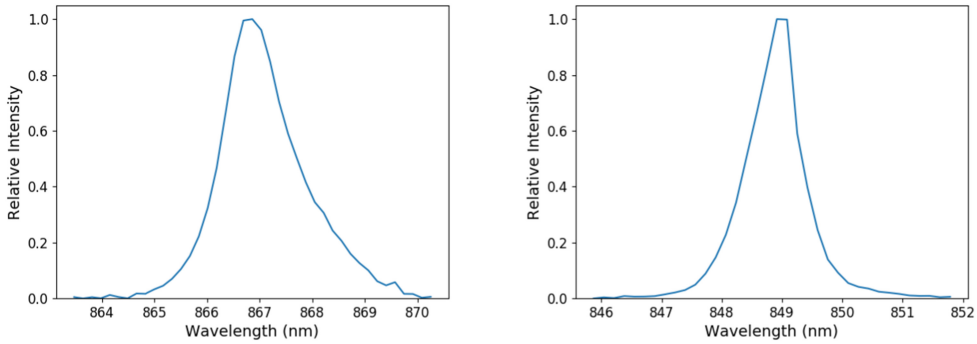
The data collection setup is shown in Figure 8, which includes (1) light source setup, (2) spectrum collection, and (3) data pre-processing. To reduce the impact of background noise, we collected data in a darkroom and fixed the relative location of the light source to the receiver. Subsequently, we adjusted the placements of the light source and receiver to make three types of photoelectric sensors: through-beam mode, retro-reflective mode, and diffuse-reflective mode. Since excessive light intensity saturates the spectrometer, we set the sensing distance for each mode at an appropriate length (10 cm–50 cm) for the same model of light sources.

A spectrometer is required to obtain a light spectrum. We used a Thorlabs Compact CCD Spectrometer CCS175 [42]. The spectrometer receiver, a cosine corrector [43], captures the light emitted by the light source. We used an optical fiber cable to transfer the optical signal to the spectrometer body. Inside, the spectrometer disperses the incident optical signal and digitizes it to create a light spectrum, which is then recorded by a laptop connected to it. Before training or evaluating Lightbox, all spectrum data must be pre-processed. This pre-processing involves cropping to focus on the spectrum of interest and normalizing its light intensity between 0 and 1. The following sections provide detailed information on the data collection process.

4.3.1 Light Sources for Evaluation. As mentioned in the Section 2, photoelectric sensors use LEDs as light sources. Therefore, instead of utilizing photoelectric sensor itself, we evaluate Lightbox using LEDs or laser modules using laser diodes, since It is very cost-efficient. Furthermore, we measured the spectrum of the light emitted from lidar [20], which is a popular photoelectric sensor, and compared it to the spectrum of the light from the laser module we used for the evaluation. Remarkably, we discovered a similarity between the two spectrums, as shown in Figure 7.

Photoelectric sensors use light sources of various wavelengths from visible to infrared. Therefore, we chose three types of LEDs and four types of laser modules with different center wavelengths. For each product, 11–20 identical models were used for experiments. As indicated in Table 1, the manufacturer provides the absolute maximum rating current, which should not be exceeded. To prevent any damage to the light sources and any changes to their spectrum, we operated the light sources within this maximum rating current.

4.3.2 Spectrum Measurement and Collection. To collect and visualize the spectrum, we used a Thorlabs spectrometer and a software package called Optical Spectrum Analyzer (OSA). The spectrometer we used has a detectable wavelength range of 500 nm–1,100 nm and a resolution of 6 px/nm, which results in a total of 3,648 pixels. This means that we can obtain a $[1 \times 3,648]$ vector depicting a spectrum from each measurement. We then configured OSA to record each snapshot of



(a) Lidar (b) Laser module

Fig. 7. Spectrum of the photoelectric sensor and laser module.

Table 1. Detailed Information of Light Sources Used in Evaluation

Identifier	Light source model	Product name	λ_C*	I_F‡	# of class
L_A	Red LED	BIWV-PR5C5T	620–660 nm	100 mA	20
L_B	875 nm IR LED	TSHA5201	875 nm	100 mA	20
L_C	880 nm IR LED	QED123	880 nm	100 mA	20
L_D	Red laser	HLM1230	650 nm	25 mA	18
L_E	IR laser	Laserland 3.5 mW laser	850 nm	350 mA	17
L_F	Green laser	AMYTA 5 mW laser	532 nm	280 mA	11
L_G	850 nm laser	1 W IR Dot Laser	850 nm	1,200mA	15

*λ_C represents the center wavelength.
‡I_F represents the absolute maximum forward current. For lasers, the maximum operating current is specified instead of the maximum continuous forward current.

the spectrum into a CSV file, with each row containing a wavelength bin paired with the matching intensity.

4.3.3 Data Preprocessing. As the effective bandwidths of the light sources are considerably narrower than that of the spectrometer, many of the wavelength bins contain near-zero intensities, which are largely meaningless. Moreover, the acquired spectrum only presents absolute intensities. As a result, unwanted bands must be cropped out, and the intensities need to be normalized.

The data preprocessing step comprised the following tasks: Initially, the effective band was extracted from the CSV files. We empirically set a threshold level to distinguish between noise and desired light signals. Then, the effective band was defined as the longest interval in which all wavelength bin intensities exceeded the threshold across all CSV data. The spectral intensities were normalized to the range of $[0, 1]$ for easier data processing. For instance, for each cropped spectrum vector of size n, $\mathbf{S} = [S_0, S_1, \ldots S_{n-1}]$, where S_i indicates the intensity of the ith wavelength bin, the normalized spectrum $\hat{\mathbf{S}}_i$ is as follows:

$$\hat{\mathbf{S}} = \left[\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_{n-1}\right], \text{where } \hat{S}_i = \frac{S_i - \min \mathbf{S}}{\max \mathbf{S} - \min \mathbf{S}}. \quad (1)$$

In the final pre-processing step, the spectrum of each light source was stored in a MongoDB database [24], which was utilized for Lightbox's training and offline testing.

4.3.4 Dataset Overview. The dataset for each light source model is summarized in Table 1. We considered 11–20 light source products of the same model for each model, and each was classified

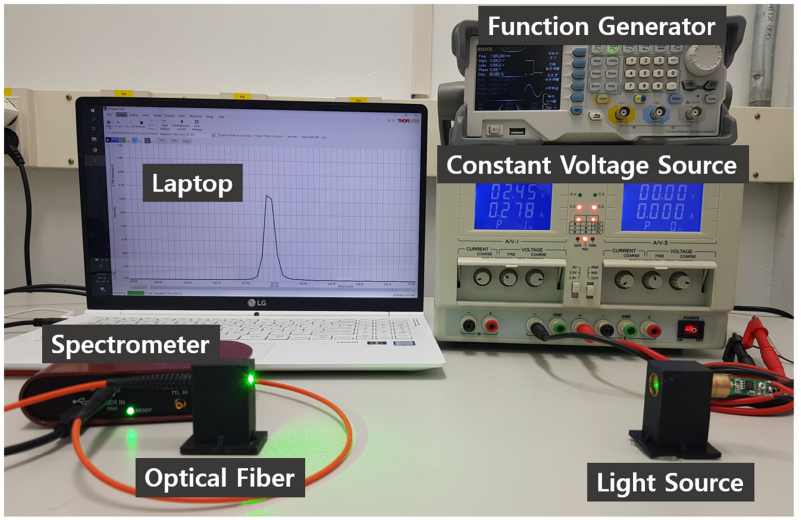


Fig. 8. Data collection setup. The light signal generated from the light source is received by the spectrometer and saved into a CSV file.

as a separate class. For each light source model, we gathered 0.3–1.2 million spectrum samples using the aforementioned experimental setup. The highest relative intensity of received light depends on the light source model, sensing mode, and distance between the light source and the receiver. To quantify the quality of the received light, we employed the optical signal-to-noise ratio (OSNR), a metric that measures the ratio of signal power to noise power within a valid bandwidth. It can be computed as $\mathrm{OSNR} = \max \mathbf{S} / \max \mathbf{S}_{noise}$, where $\max \mathbf{S}_{noise}$ is the maximum intensity of the noise.

4.4 Training Setup

4.4.1 Extractor Training Procedure. Before training the feature extractor, it is necessary to collect samples from the reference and authentic light sources using a spectrometer, as outlined in Section 4.3. To equip the feature extractor with suitable distinguishing capabilities, the spectrum of the reference light sources must be as similar as possible to that of the authentic light source.

Once the dataset was collected, the feature extractor was trained using the reference and authentic datasets. The feature extractor is a multi-class classifier that distinguishes the spectrum of light sources of the same models (authentic and reference light sources). There are no strict limitations on how accurately the multi-class classifier should be implemented. However, it must be sensitive enough to distinguish nearly identical reference light source spectrums. In this experiment, we employed a convolutional neural network (CNN) with five convolutional layers and two fully connected (FC) layers. The training operation of the feature extractor is illustrated by the upper path (blue arrows) in Figure 9

4.4.2 OC Training Procedure. This procedure follows the pre-training of the feature extractor. The authentic class dataset was used to train an OC classifier, which uses the output of the pre-trained feature extractor as input. There are no specific restrictions on the machine learning techniques utilized as long as the OC classifier can distinguish the authentic spectrum from the others. We used an OC support vector machine (OC-SVM). During the training, the feature extractor was held constant, and the OC-SVM received its input from the first FC layer. The OC training procedure is represented in the lower path of Figure 9 (green arrows).

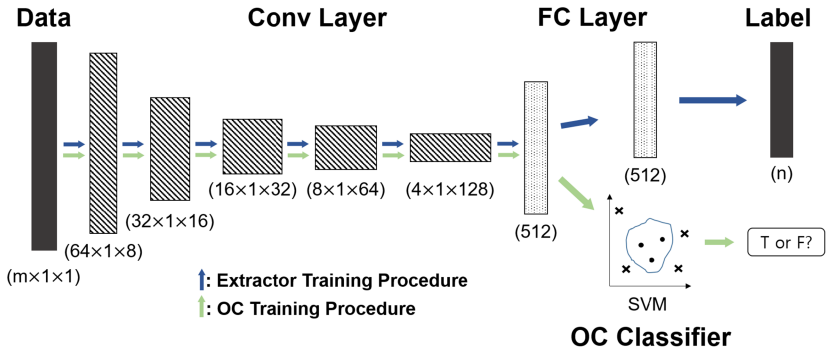


Fig. 9. Structure of the feature extractor and OC classifier. First, we trained the CNN layers through the extractor training procedure; subsequently, we trained the OC classifier using the pre-trained extractor. The output label of the OC classifier is the final detection result of Lightbox.

5 EVALUATION

We evaluate whether Lightbox meets the design goals defined in Section 4.1.

- **Q1:** How effectively can Lightbox detect sensor attacks?
  - $\mathbf{Q1}_1$ (Saturation attack): When an attacker attempts to saturate the target sensor.
  - $\mathbf{Q1}_2$ (Simple spoofing attack): When an attacker attempts to spoof the target sensor with additional light sources.
  - $\mathbf{Q1}_3$ (Sophisticated attack): When an attacker is aware of Lightbox and attempts to generate an adaptive attack signal by combining multiple light sources.
- **Q2**: Is the Lightbox compatible with various photoelectric sensors including different light source models or sensing mode? To be compatible with most photoelectric sensor applications, the system should use generic properties, instead of application-specific information.
- **Q3**: Does Lightbox work well under various environments: transmission medium, background noise, and input waveform?
- **Q4**: How does the detection performance of the Lightbox change under lumen degradation?

Note that Lightbox achieves **G2** by utilizing the spectral data, a common characteristic of all light sources, for fingerprinting, rather than relying on specific application-type features.

5.1 Performance Metrics

To evaluate Lightbox's performance, suitable evaluation metrics are required. These metrics can be divided into two categories: classwise and overall. Classwise metrics focus on evaluating Lightbox in terms of light source products (e.g., IR LED products), whereas overall metrics are used to assess Lightbox's performance with respect to light sources of a specific model (e.g., a group of IR LED products of the same model). For simplicity, each evaluation metric is represented by symbols. Let the number of classes be represented as n; the set of the spectrum samples as X; the set of samples predicted to be labeled i as $h_i(X)$ and any other label than i as $h_i^c(X)$; the set of the samples whose true label is i as $Y_i(X)$ and other than i as $Y_i^c(X)$. With these definitions, our evaluation metrics—precision, recall, F1 score, and accuracy—can be defined as follows ($|\cdot|$ denotes the number of elements of a set):

- Accuracy of ith class, $A_i = \frac{|Y_i(X) \cap h_i(X)| + |Y_i^c(X) \cap h_i^c(X)|}{|X|}$.
- Precision of ith class, $P_i = \frac{|Y_i(X) \cap h_i(X)|}{|h_i(X)|}$.

Table 2. Overall Performance of the Feature Extractor Network for Each Light Source Model

Light Source	A	P_{min}	R_{min}
L_A	**0.9975**	0.9693	0.9675
L_B	**0.9975**	0.9851	0.9900
L_C	**0.9964**	0.9639	0.9731
L_D	**0.9994**	0.9921	0.9920
L_E	**0.9999**	0.9999	0.9999
L_F	**0.9873**	0.9742	0.9749
L_G	**0.9999**	0.9987	0.9988

— Recall of ith class, $R_i = \frac{|Y_i(X) \cap h_i(X)|}{|Y_i(X)|}$.
— F1-score of ith class, $F1_i = 2 \cdot \frac{R_i \times P_i}{R_i + P_i}$.
— Minimum precision, $P_{min} = min\{P_1, P_2, \ldots, P_n\}$.
— Minimum recall, $R_{min} = min\{R_1, R_2, \ldots, R_n\}$.

A, P, R, and $F1$ notations without subscripts denote the average value of all classes.

5.2 Performance of Feature Extractor Network

Before evaluating the detection performance of Lightbox, we first assessed the feature extractor. Considering that an attacker might use a light source similar to or even of the same model as the authentic light source, Lightbox needs to be capable of distinguishing between the authentic light source and those of the same model. The feature extractor captures the differences, even among products of the same light source model. Therefore, it is necessary to check the multi-class classification performance of the feature extractor, as it directly impacts Lightbox's performance.

In this study, we adopted a CNN composed of five convolutional layers and two FC layers, as described in Section 4.4. The spectrum data for the evaluation were collected from all the light sources listed in Table 1. For each light source, the collected dataset was split into a 9:1 ratio, with each subset serving as the training and test dataset, respectively. To evaluate the performance of the feature extractor network, we used accuracy A, minimum precision P_{min}, and minimum recall R_{min} as the metrics. As a result, the accuracy was above 98% for all the light sources, as shown in Table 2. Experimental results demonstrated that the feature extractor's accuracy was high enough to distinguish even between light sources of the same product model.

5.3 Sensor Attack Evaluation (Q1&Q2)

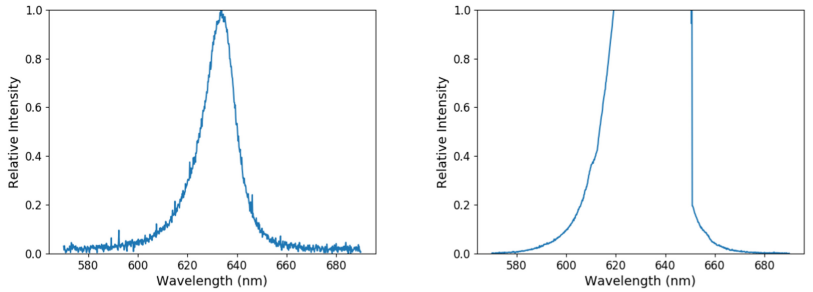
Attack methods against photoelectric sensors, as defined in Section 3, can be classified into two categories: saturation and spoofing. We evaluated attack detection performance considering (1) saturation attacks, (2) simple spoofing attacks using only one light source, and (3) sophisticated spoofing attacks that combine multiple light sources to mimic the authentic input signal.

For both saturation and spoofing attacks, the attacker can choose a light source of the same model as the authentic one or different models. Since the spectrum of a different light source model would significantly deviate from that of the authentic light source, we only considered scenarios where the attacker uses the same model as the authentic light source.

Note that evaluations were conducted with the photo-receiver and emitter in fixed positions. This setup encompasses all cases of through-beam, retro-reflective, and some instances of diffuse-reflective applications (refer to Section 2). Such setups are typical in real-world scenarios where

Table 3. Performance against the Saturation Attack

Light Source	Sensing Mode	P	R	$F1$	A	P_{min}	R_{min}
L_G	THRU	1.0000	0.9967	0.9983	**0.9990**	1.0000	0.9615
	RETRO	1.0000	0.9836	0.9917	**0.9956**	1.0000	0.9402
	DIFFUSE	1.0000	0.9946	0.9973	**0.9984**	1.0000	0.9416



(a) Spectrum of the light signal before the sensor saturation attack (b) Spectrum of the light signal after the sensor saturation attack

Fig. 10. Spectrum comparison of the L_A before and after the saturation attack.

object detection and distance measurement are crucial, such as in the manufacturing or logistics industries.

5.3.1 Saturation Attack Detection ($Q1_1$). We collected saturation attack data using light source L_G at the maximum input current specified in Table 1 to achieve the highest OSNR. Both authentic signal data and saturation attack signal data were input into Lightbox for evaluation.

Lightbox successfully detected the sensor saturation attack with an accuracy greater than 99.5%, as shown in Table 3. Impressively, the precision was 1 in all cases, and the false-positive value was zero, indicating that all saturation attack signals were correctly classified as unauthentic. Lightbox can flawlessly detect saturation attacks, because the spectrum of light received from the sensor under a saturation attack significantly deviates from the benign spectrum, as discussed in Figure 10. Consequently, we conclude that a sensor saturation attack, which aims to obstruct the sensor from receiving a benign signal, can be easily detected using Lightbox.

5.3.2 Simple Spoofing Attack Detection ($Q1_2$). To evaluate the simple spoofing attack performance, we prepared two datasets labeled True/False using the following procedure:

(i) Select the light source model and collect the data for all classes. For example, for L_E, there are 17 classes.
(ii) Exclude one class, e.g., class i, from the classes to be used as the unseen anomaly class.
(iii) Select one class, e.g., class j, from the remaining classes to be used as True data (authentic), and use the other remaining classes as False data (anomalies for training).
(iv) Train Lightbox with classes except for j and a fraction of i (authentic data for testing), and test it against the remaining i and j (anomaly data for testing).
(v) Repeat (ii)–(iv) for every (i, j) pair for cross-validation.

Through-beam Application

In this section, we evaluated Lightbox's overall performance on a through-beam sensor application. We set up the experiment with the light source and spectrometer receiver facing

Table 4. Performance of Lightbox for Each Sensing Mode

Sensing Mode	Light Source		P	R	$F1$	A	P_{min}	R_{min}
THRU	L_A		0.9988	0.9624	0.9802	**0.9792**	0.9985	0.8769
	L_B		0.9987	0.9470	0.9722	**0.9708**	0.9980	0.9007
	L_C		0.9976	0.9492	0.9728	**0.9716**	0.9953	0.8904
	L_D		0.9990	0.9672	0.999	**0.9819**	0.9988	0.9007
	L_E		0.9993	0.9958	0.9976	**0.9975**	0.9984	0.9537
	L_F		0.9991	0.9721	0.9857	**0.9851**	0.9985	0.9376
	L_G		0.9993	0.9967	0.9980	**0.9980**	0.9985	0.9615
RETRO	L_E		0.9981	0.9784	0.9881	**0.9878**	0.9898	0.9096
	L_F		0.9988	0.8388	0.9119	**0.8905**	0.9973	0.6146
	L_G		0.9995	0.9836	0.9915	**0.9912**	0.9989	0.9402
DIFFUSE	L_G	CARBON BOX	0.9992	0.9946	0.9969	**0.9968**	0.9989	0.9416
		PLASTIC PLATE	0.9994	0.9842	0.9917	**0.9914**	0.9978	0.9242
		IRON PLATE	0.9992	0.9861	0.9926	**0.9924**	0.9982	0.9373
		CONCRETE WALL	0.9991	0.9672	0.9829	**0.9822**	0.9985	0.9177

Table 5. OSNR of Each Light Source in Three Sensing Modes at a Sensing Distance of 30 cm

Sensing Mode	$OSNR_A$	$OSNR_B$	$OSNR_C$	$OSNR_D$	$OSNR_E$	$OSNR_F$	$OSNR_G$
THRU	**2.3890**	**3.0625**	**6.4279**	**258.0402**	**334.2079***	**318.3503***	**227.2429***
RETRO	0.8381	1.0317	0.9812	0.9281	**2.2549**	**1.7889**	**227.2429***
DIFFUSE	0.9290	1.0114	0.9826	1.0426	0.9817	0.9741	**3.6957**

*The signal power exceeded the maximum threshold of the spectrometer can receive.

each other at a fixed sensing distance, with no objects in between, to emulate a through-beam application. The influence of the transmission medium and distance is discussed later in this section.

The experiment results are presented in Table 4. Lightbox showed consistently high performance for all seven light sources. The accuracy was higher than 97% for all the light sources, and the accuracy of the laser was higher than those of LEDs L_A, L_B, and L_C. The results showed that Lightbox could distinguish between the authentic light source and other light sources of the same model for the through-beam application.

Retro-reflective Application

We evaluated the performance of Lightbox on retro-reflective sensor applications. We first measured the OSNR for each light source and sensing mode at a sensing distance of 30 cm. The results are listed in Table 5. The light sources whose optical power was low (i.e., OSNR was lower than a threshold) were excluded from the retro-reflective experiments. If the intensity of the light source is too weak, then performance of the Lightbox decreases due to the low signal quality as shown in Section 7.1. Therefore, we set a threshold value for the OSNR and use only the light source that exceeds the threshold value for the evaluation. We experimentally found that the performance of the Lightbox decreases under 0.7 when OSNR of the light source is lower than 1.5, so we set the threshold to 1.5. L_A, L_B, L_C, and L_D were excluded from this evaluation. We placed the light source and the receiver of the spectrometer on the same side and placed a retro-reflector on the opposite side at a fixed sensing distance for each light source to construct the retro-reflective application environment. We collected training data at a sensing distance of 30 cm and evaluated the attack detection accuracy.

The experimental results are given in Table 4. The overall detection accuracies for the retro-reflective application using light sources L_E, L_F, and L_G were 98.78%, 89.05%, and 99.12%, respectively. Note that the reason why the detection accuracy of L_F was significantly low is probably that the measured OSNR of L_F was 1.789, which was much lower than other light sources. The relationship between OSNR and attack detection accuracy is discussed more in Section 7.1.

Diffuse-reflective Application

We evaluated the performance of Lightbox on diffuse-reflective sensor applications. We placed the light source and the receiver of the spectrometer on the same side and placed a reflecting object on the opposite side at a distance of 20 cm to construct the diffuse-reflective application environment. Moreover, we used four types of reflecting objects: carbon box, plastic plate, iron plate, and concrete wall. We used L_G for the experiment and other light sources whose OSNR was lower than the threshold (1.5) were excluded from the diffuse-reflective experiments.

The experiment results are presented in Table 4. The detection accuracy for the diffuse-reflective application using L_G was consistently higher than 98% regardless of the type of reflecting object, indicating that Lightbox could effectively detect sensor attacks, even for diffuse–reflective sensor applications.

For the rest of the evaluations, we used L_G as a light source, since it has sufficient power to support all sensing modes. Notably, we used a carbon box as a reflecting object in the diffuse-reflective mode for further experiments in Q3 and Q4.

5.3.3 Sophisticated Spoofing Attack Detection ($Q1_3$). While a simple spoofing attacker attempts to spoof the target device by using only one light source, we now consider a sophisticated spoofing attacker who is aware of Lightbox and uses multiple light sources. A sophisticated attacker generates malicious input signals in the following steps: (1) the attacker eavesdrops on the spectrum of the authentic light source and (2) combines the multiple light source signals to mimic the received signal. An attacker can use multiple light sources from various light source models. However, when the number of light sources is fixed, mimicking an authentic spectrum is easier using light sources of the same model as the authentic light source, rather than using light sources from various models. Therefore, we assume that the attacker uses multiple light sources of the same model as the authentic light source.

Eavesdropping on Authentic Spectrum Data

For a sophisticated spoofing attack, the attacker should know the spectrum data of the authentic light source, since Lightbox detects sensor attacks based on the spectrum shape of the received signals. Since we assume the attacker has no direct access to the target (Section 3.2), the attacker cannot receive the spectrum by tampering with the target device's hardware and software. Therefore, the attacker should receive the authentic light source signal close to the target system, which causes a relative position difference between the authentic light source and receiver.

Here, we evaluated Lightbox's performance when a sophisticated attacker eavesdrops on the spectrum data close to the victim's photo-receiver and replicates it for an attack. More specifically, we gathered spectrum data from the light source at slightly altered positions to evaluate whether Lightbox could distinguish these minor changes as indications of a malicious attack. The spectrum data were collected for two cases: (1) with the angle of incidence at 0/15 degrees while the sensing distance was fixed at 50 cm and (2) with a sensing distance of 50 cm/30 cm while the angle of incidence was fixed at zero degrees. First, we trained Lightbox with the data collected at an angle of incidence at zero degrees and a sensing distance of 50 cm. Second, we tested the system by using data collected at various angles or distances. The light source used for the experiment is L_G.

Table 6. Detection Performance over the Different Relative Positions

Light Source	Sensing Mode	$\Delta\,\theta^*$	$\Delta\,d^\ddagger$	P	R	$F1$	A	P_{min}	R_{min}
L_G	THRU	0	20	**0.2060**	1.0000	0.3416	0.6030	0.0000	1.0000
		15	0	**0.0942**	1.0000	0.1722	0.5451	0.0000	1.0000
	RETRO	0	20	**0.4783**	0.7626	0.5879	0.7223	0.0000	0.0000
		15	0	**0.3744**	0.6021	0.4617	0.6680	0.0000	0.0000
	DIFFUSE	0	20	**0.4827**	1.0000	0.6551	0.7204	0.0000	1.0000
		15	0	**0.2797**	1.0000	0.4371	0.6221	0.0000	1.0000

*$\Delta\,\theta$ represents the difference in angle of incidence between training and test dataset.
‡$\Delta\,d$ represents the sensing distance difference between training and test dataset.

The experiment results are shown in Table 6. The precision of the evaluation did not exceed 48%, which indicates that the spectrum collected at different distances or angles were different compared to that at the original position. It concludes that even a sophisticated attacker, capable of perfectly replicating the eavesdropped light signal, is hard to bypass Lightbox. The result can be explained by the following reason: The difference in the receiver position causes differences in the path of light, which in turn alters the phase of the light, leading to a spectral change.

Combining Multiple Light Sources

We evaluated how similar an attacker can generate a sophisticated attack spectrum by combining multiple light source signals. The sophisticated spoofing attack consists of two steps: finding the optimal ratio of multiple light sources and injecting the attack signal based on the optimal ratio.

Attack Optimization Evaluation When two light signals are received simultaneously, the spectrum of the received light is the sum of the spectrum of each light signal. The attacker's approach of combining multiple spectra for an attack can be expressed as a weighted sum of the spectrum of each light signal. Therefore, the attacker can collect the spectrum of each light source and find the optimal ratio to generate the desired spectrum in advance.

To evaluate the detection performance against the sophisticated attack, we first trained Lightbox with 50 light sources of the same model. We assumed that the attacker owns the spectrum of the authentic light source. The attacker can determine the number of light sources to be used in the attack and find the optimal combination to mimic the authentic spectrum. For the optimization, we used the L-BFGS-B method [55] to minimize the norm 1 distance defined. We used up to 50 light sources for the attack evaluation to check the relationship between the detection performance of Lightbox and the number of attack light sources.

The evaluation results are illustrated in Figure 11. The attack accuracy is calculated by (# of optimized spectrum data classified as authentic data)/(total # of optimized spectrum data). The accuracy did not exceed 20%, even though the attacker used 50 light sources for the optimization. We also evaluated the norms 0, 2, and inf in the optimization algorithm and observed that the attack success rate was the highest when the attacker used norm 1 for the optimization. In conclusion, we experimentally confirmed that Lightbox could effectively detect sophisticated sensor attacks even if the attacker used 50 light sources to mimic the authentic spectrum.

Combining Signals in Real World When the attacker generates the attack signal in the real world, it may differ from the desired spectrum, since noise exists when the spectrum of each light source is received. Therefore, we evaluated the difference between the real-world attack signal and the linear summation of each intensity of the light sources' spectrum. To evaluate the spectrum difference, we first selected two light sources. Second, we collected the spectrum of each light

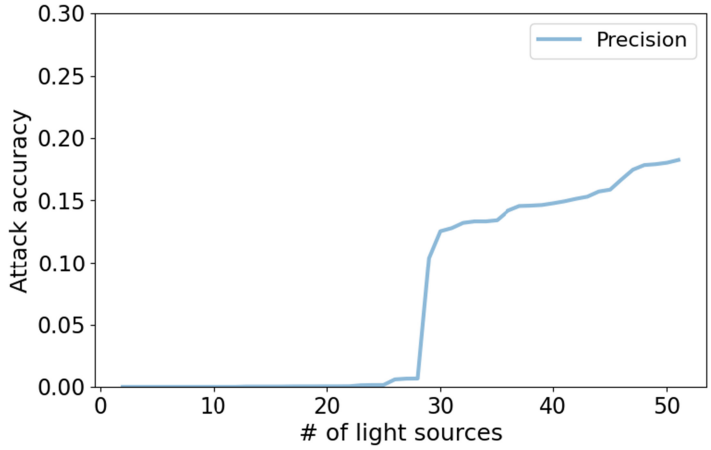


Fig. 11. Attack accuracy over the number of the light sources used for the attack.

source and defined the sum of the spectrum as the spectrum of the authentic light source. Then, Lightbox was trained with the defined authentic spectrum data. Subsequently, we collected the spectrum when two light sources were turned on simultaneously and evaluated the precision of the real-world signal. We performed cross-validation by repeating the whole step for all the possible light-source combinations of two numbers.

As a result, the precision for the real-world signal was less than 0.3945, indicating that the real-world combined signal is identified as an authentic signal with a probability of 0.3945, which is significantly low. The result shows that Lightbox does not classify the spectrum of the combined signal as being the same as the sum of the spectra of each light source. Therefore, an attacker cannot generate a spectrum that matches the one pre-calculated in the optimization stage, due to the noise from real-world injection.

In summary, we confirmed that a sophisticated attacker cannot replicate the exact spectrum of the authentic signal, and the success rate of an attack that merges multiple light sources was not high enough to bypass Lightbox.

5.4 Detection Performance against Various Environment Variables (Q3)

Photoelectric sensors are used with various factors, such as (1) background noise signals, (2) input current, and (3) the medium between the transmitter and the receiver. We evaluated the detection performance of Lightbox under the influence of different environmental factors.

5.4.1 Impact of Background Noise. In real-world applications, photoelectric sensors are not placed in a darkroom, as was the case in the experimental environment. Several external light sources, such as fluorescent lights, coexist in the background. Therefore, we evaluated the influence of background noise on the performance of Lightbox. We constructed the environment with background noise by receiving light in a bright room with fluorescent lights turned on. The illuminance values of the darkroom and the noisy environment with fluorescent lights are 2 lx and 793 lx, respectively. The data collected in the darkroom are denoted as D_D, and the noisy environment with fluorescent lights as D_N. We trained Lightbox with D_D and tested the overall performance with D_N.

The resulting performance of Lightbox is listed in Table 7. The detection performance of Lightbox in a noisy environment was consistently high and did not decrease. Based on the result, we can conclude that background noise does not influence the performance of Lightbox.

Table 7. Performance over Different Environment Noises

Light Source	Sensing Mode	Train	Test	P	R	$F1$	A	P_{min}	R_{min}
L_G	THRU	D_D	D_D	0.9993	0.9967	0.9980	0.9980	0.9985	0.9615
		D_D	D_N	0.9994	0.9998	0.9996	**0.9996**	0.9987	0.9987
	RETRO	D_D	D_D	0.9995	0.9836	0.9915	0.9912	0.9989	0.9402
		D_D	D_N	0.9989	0.9937	0.9963	**0.9961**	0.9983	0.9180
	DIFFUSE	D_D	D_D	0.9992	0.9946	0.9969	0.9968	0.9989	0.9416
		D_D	D_N	0.9990	0.9857	0.9923	**0.9920**	0.9985	0.9241

Table 8. Performance with Pulsed Light Source

Light Source	Sensing Mode	P	R	$F1$	A	P_{min}	R_{min}
L_G	THRU	0.9947	0.9998	0.9972	**0.9972**	0.9863	0.9990
	RETRO	0.9987	0.9838	0.9911	**0.9910**	0.9960	0.9497
	DIFFUSE	0.9987	0.9878	0.9932	**0.9931**	0.9977	0.9712

5.4.2 Impact of Square Waveform. All the previous evaluations were conducted based on the assumption that the photoelectric sensor receives constant light signals. However, some photoelectric sensors, such as OBSDs and drop sensors, utilize pulsed light. Therefore, we evaluated the detection performance of Lightbox when input signals were generated from pulsed light sources. We generated an input current as a square waveform with a frequency of 1 Hz and 50% duty cycle and used it for data collection. The data collected with square input current is denoted as D_S. We trained Lightbox with the D_S training dataset and tested the overall performance with the D_S test dataset.

Table 8 shows the experimental results. The accuracy of the pulse input signal was higher than 99%. The result showed that Lightbox could also be applied to pulse input waveform applications as long as it was trained by the data collected using the same type of current used in photoelectric sensor applications.

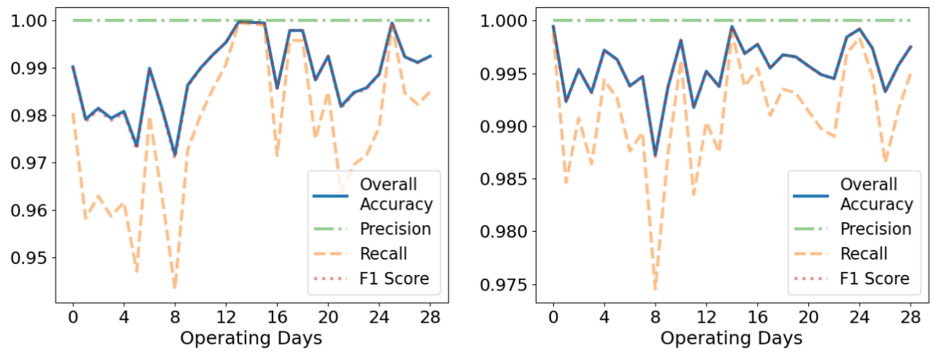
5.4.3 Impact of the Transmission Medium. In real-world applications, photoelectric sensors are used in various transmission media. For example, photoelectric sensors can be used to detect liquid in a bottle or to detect smoke [2, 10]. Therefore, we evaluated the capacity of Lightbox to classify the authentic light source even with a non-air transmission medium. We conducted the experiment using the spectrum data transmitted with various types of media between the light source and receiver. We selected four types of media, namely, clean air, a 3 cm thick transparent acrylic plate, 20 cm wide transparent glass bottle filled with pure water, and smoke with 0.8 light intensity attenuation [23]. Subsequently, we evaluated if Lightbox can distinguish the authentic light source when the four different transmission media were located between the light source and the receiver of the spectrometer. Note that the trained models were evaluated using the test dataset collected in the same environment as the training dataset (the training data and testing data were collected in the same medium).

The experiment results are shown in Table 9. The results indicate that regardless of the medium between the light source and receiver, Lightbox can recognize the authentic light source with an accuracy higher than 99%. However, when we tested Lightbox with the dataset collected with different media from the training dataset, the performance of Lightbox significantly decreased.

Although Lightbox does not operate well for other media, it can be applied with consistently high accuracy to the types of photoelectric sensors where light must pass through several media,

Table 9. Performance over Different Media

Light Source	Sensing Mode	Medium	P	R	$F1$	A	P_{min}	R_{min}
L_G	THRU	Air	0.9993	0.9967	0.9980	**0.9980**	0.9985	0.9615
		Acrylic	0.9986	0.9880	0.9933	**0.9926**	0.9922	0.8658
		Smoke	0.9991	0.9904	0.9947	**0.9944**	0.9984	0.9154
		Water	0.9976	0.9890	0.9933	**0.9929**	0.9792	0.9038
	RETRO	Air	0.9995	0.9836	0.9915	**0.9912**	0.9989	0.9402
		Acrylic	0.9993	0.9789	0.9890	**0.9883**	0.9988	0.9129
		Smoke	0.9989	0.9885	0.9937	**0.9934**	0.9980	0.9248
		Water	0.9990	0.9842	0.9915	**0.9911**	0.9983	0.9216
	DIFFUSE	Air	0.9992	0.9946	0.9969	**0.9968**	0.9989	0.9416
		Acrylic	0.9991	0.9916	0.9953	**0.9950**	0.9987	0.9057
		Smoke	0.9990	0.9946	0.9968	**0.9967**	0.9985	0.9367
		Water	0.9990	0.9848	0.9918	**0.9915**	0.9984	0.9330



(a) L_A Lightbox (b) L_G Lightbox

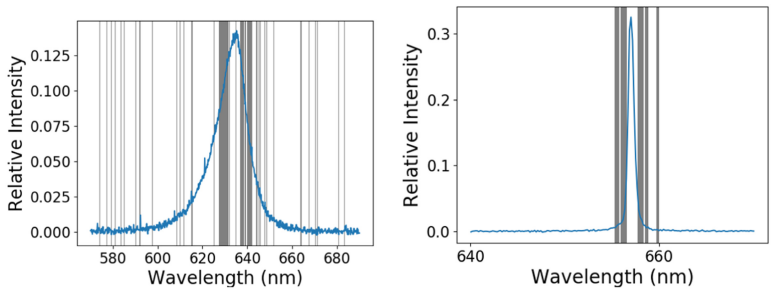
Fig. 12. Performance (through-beam) over four weeks.

such as infusion pumps, through the comprehensive use of the trained model for each medium in the form of a decision tree, since the accuracy of the system for each medium is higher than 99%.

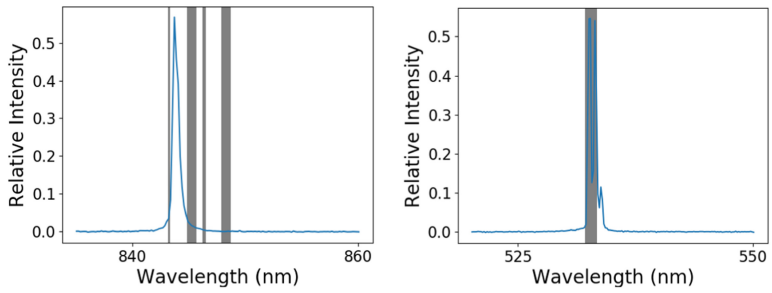
5.5 Longer-term Consistency (Q4)

As is common knowledge, LEDs and lasers have limited lifetimes; thus, lumen degradation and color shifts may occur over time. Therefore, we evaluated whether the pre-trained Lightbox could consistently identify the authentic light source over time. To evaluate the performance of Lightbox over time, we collected the spectrum data from light sources L_A and L_G for one day to create the training dataset. We then collected the spectrum data of the authentic light source every day over four weeks. After the data collection, we tested the system performance over time to determine the authenticity of the incoming light for L_A and L_G. Note that the through-beam sensing mode was used for this experiment, and during our long-term consistency experiment, we collected data with the light source continuously on, without switching it on and off.

The evaluation results of the longer-term performance are shown in Figure 12. For both light sources, the accuracy did not decrease below 97%. The results reveal that the performance of the pre-trained Lightbox is consistently maintained over a minimum of four weeks.



(a) LIME result of L_A (b) LIME result of L_D



(c) LIME result of L_E (d) LIME result of L_F

Fig. 13. LIME interpretations for each type of light source; the blue lines indicate the spectrum of the light source, and the wavelength segments with significant influence on the classification results are indicated by gray shades.

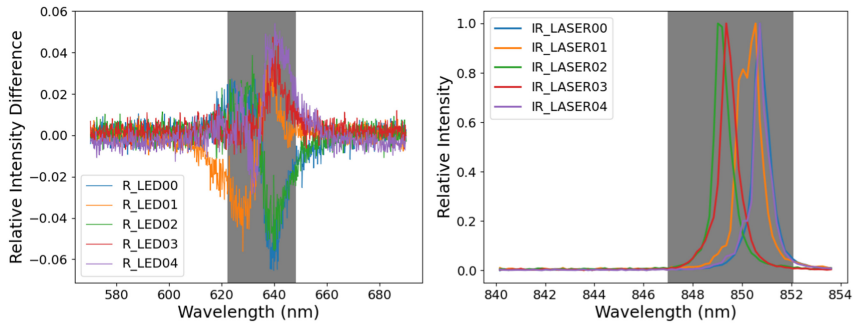
6 INTERPRETABILITY

Although our experimental results indicate the effectiveness of Lightbox, it is still unclear how the system determines the authenticity of the incoming signal. To address this issue, we attempted to interpret the decisions made by Lightbox using two methods, namely, LIME [31] and feature selection of the spectrum data.

6.1 Interpretation Using LIME

LIME [31] is a popular method for interpreting decisions made by black-box machine learning models. It perturbs a sample to create a set of neighbors and acquires corresponding outputs for the sample and neighbors by querying the black-box model of interest. These perturbed neighbors and their corresponding decisions are then used to train an interpretable surrogate model, such as linear regression or a decision tree, in a supervised manner. The loss function used to train the surrogate is weighted based on the proximity of the perturbed neighbors to the sample of interest to ensure that the black-box model is locally approximated by the decision boundary. Following these stages, a locally interpretable approximation of the black-box model is created, and the decision boundary of the surrogate is used to interpret the sample of interest.

We applied LIME to analyze the relationship between an input spectrum and the corresponding decisions of Lightbox. Specifically, we randomly selected 1,000 samples regardless of the class and used LIME on the feature extractor to determine the influence of each small interval of the data. With the LIME results for all selected samples from each class, we obtained numerical values indicating how much each spectral segment influences the classification results. The segments with the top 20% influence are highlighted in gray. Figure 13 presents the results of



(a) Spectrum difference of the L_A (b) Spectrum of the L_E

Fig. 14. Spectrum data for feature selection evaluation.

the interpretation, which show that the segments of the wavelengths with the most influence are concentrated near the peaks. This is plausible, because wavelengths far from the peaks do not correspond to light sources; instead, they correspond to ambient noise, which is intuitively meaningless for the classification of light sources.

6.2 Feature Selection of the Spectrum

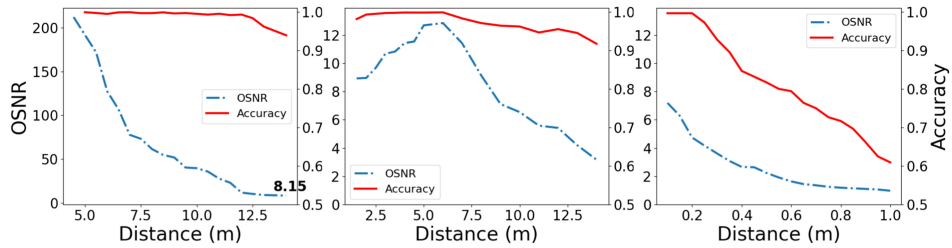
The result obtained via LIME demonstrates that spectral intensities near the spectral peak have a significant influence on the classification of each light source. For further evaluation, we partitioned each spectrum into small segments and trained Lightbox using these segments to identify which part of the spectrum most significantly influences Lightbox. To do this, we used L_A and L_E as light sources and truncated the existing $[1 \times N]$ spectral vector into 8–9 small vector segments. Specifically, L_A was divided into 9 $[1 \times 80]$ vectors and L_E into 8 $[1 \times 10]$ vectors. Following the segmentation, one of the segments was selected to train and test Lightbox. The overall performances corresponding to each selection are presented in Table 10. For the case of L_A, the overall performance of Lightbox was high with input intervals near 400, which corresponds to the peak of the spectrum. The performance decreased as the input data moved away from the peak of the spectrum. Moreover, when Lightbox was trained using the interval 320–480 near the spectrum peak (indicated by the gray shade in Figure 14(a)), the accuracy was 93.81%. Similarly, for the case of L_E, the overall performance of Lightbox improved as the input data approached index 50, which corresponds to the peak of the spectrum (the gray shade in Figure 14(b)). The performance degraded as it moved away from the peak of the spectrum. When Lightbox was trained using the interval 40–70, the accuracy was 98.26%. Note that the data in Figure 14(a) are measured by the difference between the normalized spectrum of each LED product and the average of the normalized spectrum of all LED products.

Overall, the performance of Lightbox is higher than 85% for all but R_{min}, even though it is trained using only a small portion of data in the middle. The low performance for R_{min} is explained in Figure 14. Figure 14(a) shows that the spectrum differs from the average intensity values of the 5 L_A products, and the gray-colored part represents the interval [320, 480]. As shown in Figure 14(a), the gray-colored part does not fully cover the spectrum difference, leading to performance degradation for certain classes. Similarly, for L_E, some parts of the spectrum are not covered by the gray section, as shown in Figure 14(b).

The results demonstrate that the data near the peak of the spectrum is the most important factor in classifying light sources. We also note that the difference in model accuracy between using only the corresponding section data and using the entire part is not significant.

Table 10. Evaluation Result of Feature Selection

Light Source	Interval	P	R	$F1$	A	P_{min}	R_{min}
L_A	0–80	0.9952	0.5014	0.6669	0.5028	0.9943	0.5006
	80–160	0.9969	0.5015	0.6673	0.5030	0.9963	0.5006
	160–240	0.9832	0.5082	0.6700	0.5156	0.9817	0.5024
	240–320	0.9852	0.5160	0.6773	0.5299	0.9839	0.5027
	320–400	0.9982	0.7018	0.8242	**0.7752**	0.9979	0.6166
	400–480	0.9987	0.7546	0.8597	**0.8355**	0.9984	0.6883
	480–560	0.9969	0.5267	0.6892	0.5479	0.9961	0.5003
	560–640	0.9969	0.5013	0.6672	0.5027	0.9964	0.5005
	640–720	0.9963	0.5011	0.6668	0.5022	0.9955	0.5008
	320–480	0.9984	0.8554	0.9214	**0.9120**	0.9982	0.7681
L_E	0–10	0.9979	0.5227	0.6861	0.5263	0.9972	0.5002
	10–20	0.9979	0.5219	0.6854	0.5288	0.9977	0.5003
	20–30	0.9981	0.5668	0.7230	0.5743	0.9978	0.5003
	30–40	0.9981	0.5400	0.7008	0.5481	0.9969	0.5003
	40–50	0.9982	0.6977	0.8213	**0.7493**	0.9958	0.5001
	50–60	0.9991	0.9254	0.9608	**0.9341**	0.9972	0.5103
	60–70	0.9987	0.7356	0.8472	**0.7844**	0.9958	0.4993
	70–80	0.9979	0.5554	0.7136	0.5699	0.9964	0.4999
	40–70	0.9991	0.9779	0.9883	**0.9826**	0.9975	0.6426



(a) Through-beam (b) Retro-reflective (c) Diffuse-reflective

Fig. 15. OSNR and accuracy as a function of distance.

7 DISCUSSION

7.1 OSNR and Accuracy with Respect to Distance

As photoelectric sensors detect objects based on differences in the quantity of received light, the OSNR of the received light is crucial. In this section, we measure both the accuracy and OSNR of Lightbox with respect to varying sensing distances in three sensing modes. We compare the performances of Lightbox in the three modes directly, using L_G as the light source.

The results are shown in Figure 15. In all sensing modes, the accuracy of Lightbox decreases with an increase in sensing distance. In the through-beam sensing mode, the accuracy degrades to 93% at a sensing distance of 14 m. Although this accuracy is relatively low, it is still sufficient to detect most attacks. The OSNR at a sensing distance of 14 m is 8.15. In the diffuse-reflective mode, accuracy deteriorates as the OSNR decreases below 5, reaching a final accuracy of 60%. Conversely, in the retro-reflective mode, the OSNR slightly increases as the sensing distance is increased to 5 m, which aligns with the trend observed in the excess gain graph depicted in Figure 2. Consequently,

the accuracy also slightly increases as the sensing distance is increased to 5 m, then decreases as the sensing distance is further increased. These results demonstrate a positive correlation between the accuracy of Lightbox and the OSNR of the received light, indicating that a sufficiently large OSNR is required for stable system accuracy.

7.2 Practicality

This section demonstrates the practicality of Lightbox. First, we (1) suggest an application scenario for Lightbox. Then, we discuss practical considerations in detail: (2) reducing training efforts and (3) implementing Lightbox in embedded systems.

7.2.1 Application Scenario. Consider industries such as surveillance, medical devices, autonomous driving, industrial safety, and theme park safety. These sectors extensively use photoelectric sensors for tasks such as object detection and production line monitoring and could benefit from using Lightbox for detecting sensor attacks. For instance, Lightbox can enhance safety on production lines by detecting anomalies in item sizes or spotting errors such as misaligned bottle caps. Similarly, in automated warehouses, it can assist in object sensing to improve operational efficiency and safety.

However, there are two main considerations: First, clients may face challenges in preparing sufficient reference data for model training before using Lightbox. Second, Lightbox needs to be lightweight enough to operate effectively in an embedded system. We address these two practical considerations in Sections 7.2.2 and 7.2.3.

7.2.2 Reducing Training Efforts of Lightbox. As described in Section 4, the spectra of the authentic light source are required for the training of the feature extractor and the OC classifier. In cases of mass deployment, this necessitates the training of the entire Lightbox by the manufacturer for every site or by the user using a large number of reference datasets provided by the manufacturer.

To minimize this process efficiently, we first evaluated the model's performance using a reduced amount of training data. We found that the model still achieved high performance when trained with only about 100,000 samples. These samples can be gathered in just 5 to 10 minutes for each product.

Moreover, we estimated the accuracy and training duration of Lightbox using data solely from the authentic light source to fine-tune the feature extractor, instead of adopting the original training method [29]. In this case, the manufacturer only needs to mount the initial feature extractor trained with the reference dataset, thereby avoiding any device-specific training. Subsequently, the user can fine-tune the model without the reference dataset.

The feature extractor is first trained using only the reference dataset. Then, the data of the authentic class are utilized to train only two layers of the network, namely, the final convolutional layer and the FC layer connected to it. Finally, the OC classifier is trained using the fine-tuned feature extractor.

In Table 11, Lightbox trained via the original method is denoted as M_N, and that trained using the fine-tuning method is denoted as M_F. The results show that irrespective of the sensing modes, the accuracy of Lightbox trained via fine-tuning was slightly lower than that trained using the original model. However, fine-tuning effectively reduced the training duration. Furthermore, the original training method required users to access a 4.7 GB reference dataset consisting of the spectrum data from 50 reference light sources. In contrast, the model size required for fine-tuning was merely 180 MB. Therefore, fine-tuning can reduce both data size and training duration with negligible accuracy degradation.

7.2.3 Implementation Using a Single Board Computer. During the evaluation process described in Section 5, Lightbox was trained and tested on a server with sufficient computational capacity.

Table 11. Performance of the Fine-tuned Lightbox for a Through-beam Application

Sensing Mode	Training	P	R	$F1$	A	P_{min}	R_{min}	Training duration (s)	
								Server	Jetson
THRU	M_N	0.9993	0.9967	0.9980	0.9980	0.9985	0.9615	42	234
	M_F	0.9992	0.9835	0.9913	**0.9904**	0.9988	0.8666	11	186
RETRO	M_N	0.9995	0.9836	0.9915	0.9912	0.9989	0.9402	44	265
	M_F	0.9990	0.9919	0.9954	**0.9952**	0.9981	0.9253	16	209
DIFFUSE	M_N	0.9990	0.9848	0.9918	0.9915	0.9984	0.9330	45	242
	M_F	0.9989	0.9827	0.9907	**0.9903**	0.9983	0.9210	12	196

The training duration of Lightbox depends on the hardware used. Although detection accuracy remains constant given sufficient time, excessively long training periods reduce the applicability of Lightbox. To verify this, Lightbox was implemented on a state-of-the-art single-board computer, Jetson TX2, which runs on a quad-core ARM Cortex-A CPU and 8 GB RAM. The training duration required by Lightbox via fine-tuning on the server and Jetson TX2 is presented in Table 11. This confirms that although the training duration required for Jetson TX2 is longer than that of the server, it is still an acceptable duration. As a result, it confirms that Lightbox can be trained and operated within a short period, even in a practical scenario with a single-board computer.

Operation of Lightbox with a Real-world Application We considered applying a simple through-beam sensor using Jetson TX2 and an STS-NIR microspectrometer [25] to instantiate an end-to-end application scenario. The sensor received incoming light signals through the spectrometer and indicated the authenticity of their sources using green and red LEDs. The sensing distance was set to 3 m, and L_G was used as the light source. Note that training and test data were collected using an STS-NIR microspectrometer.

The following three scenarios were tested: (1) involving only authentic incoming signals; (2) involving only adversarial incoming signals; and (3) involving a combination of adversarial and authentic incoming signals of equal intensity. The distance between the adversarial light source and the target sensor was considered to be equal to the sensing distance. Furthermore, the adversarial light source maintained an angle of 10 degrees with the straight line from the authentic light source to the target sensor in scenarios (2) and (3). Based on the assumption that the attacker is incapable of directly damaging the victim's hardware, the aforementioned placement of light sources is a plausible attack scenario. The target sensor reacted sensitively to changes in the incoming light, and Lightbox was observed to respond appropriately in all the test scenarios. This confirms the practicality of Lightbox in a real-world application. A demonstrative video of our experiment is available at https://youtu.be/_KSq3EvUwt4.

7.3 Limitation

Robustness with Respect to Number of Light Sources One of the traditional challenges in fingerprinting research is the decline in model accuracy as the size of the test dataset increases. Although we could not utilize a vast number of light sources for our evaluations, we demonstrated that Lightbox can effectively detect sensor attacks with a relatively small training dataset. Specifically, we conducted two experiments with 100 light sources: (1) measuring the accuracy corresponding to a large reference dataset and (2) measuring the accuracy when a large number of adversarial light sources were involved. To evaluate the accuracy relative to a large reference dataset, we used 100 light sources of the L_A model. From these, 20 light sources were randomly selected to act as adversarial light sources. Finally, one of the remaining 80 light sources was set to be the authentic source, and the others were taken to form the reference dataset. As presented

Table 12. Performance of Lightbox for 100 Light Sources

Light Source	Reference #	Anomaly #	P	R	$F1$	A	P_{min}	R_{min}
L_A	80	20	0.9992	0.9681	0.9834	**0.9821**	0.9940	0.8026
	50	50	0.9991	0.9707	0.9847	**0.9839**	0.9950	0.8586
	40	50	0.9992	0.9763	0.9876	**0.9870**	0.9984	0.8136
	30	50	0.9991	0.9717	0.9852	**0.9839**	0.9982	0.8018
	20	50	0.9992	0.9872	0.9931	**0.9929**	0.9980	0.9804
	10	50	0.9989	0.9444	0.9709	**0.9673**	0.9976	0.9009

in Table 12, the detection accuracy is 98.21%, which is similar to the accuracy of 97.92% for L_A light sources, as mentioned in Table 4.

To evaluate the detection accuracy against a large number of adversarial light sources, we initially selected 50 light sources among 100 light sources of the L_A model. Subsequently, groups of 10, 20, 30, 40, and 50 light sources were selected from the remaining 50 light sources. Within each group, one light source was chosen to be the authentic light source, and the others were used as reference light sources. We repeated this process twice to evaluate the impact of reference dataset size. As shown in Table 12, even when the reference dataset size decreases, the accuracy does not decrease significantly.

In conclusion, we demonstrated that Lightbox can effectively detect sensor attacks even when dealing with a large number of test datasets. Further studies on the robustness of Lightbox concerning the number of light sources are left for future work.

Robustness with Respect to Variation of Spectrometer In this study, the Compact CCD Spectrometer and the STS-NIR microspectrometer were used for experiments. However, the spectrum's shape may vary, depending on the spectrometer model. Since Lightbox uses the spectrum as input data, the spectrometer model might affect Lightbox's performance. We leave this as a topic for future work.

Longer-term Consistency Although we have verified the long-term reliability of Lightbox over a period of four weeks in Section 5.5, the typical lifetime of photoelectric sensors ranges from several months to years, depending on their application. For such longer-term use, a test period of four weeks may be insufficient. For instance, light sources may undergo hardware aging, which can gradually alter their spectra, leading to potential misclassification of authentic light sources as anomalies by Lightbox. However, we propose that such issues can be addressed by adopting partial online training for Lightbox, a topic we leave for future work. Lightbox comprises a pre-trained feature extractor and a finely tuned one-class classifier. By gradually updating the finely tuned one-class classifier while preserving the pre-trained part, Lightbox can avoid the degradation of the authentic light source while limiting the overhead of retraining.

8 RELATED WORK

Attacks against Through-Beam and Retro-reflective Sensors As evaluated in Section 5.3, Lightbox can detect sensor attacks on through-beam, retro-reflective, and diffuse-reflective sensors. Previous studies have introduced real-world attacks on these three types of sensors. Park et al. [27] demonstrated critical vulnerabilities in two models of a medical infusion pump used for intravenous therapy. These pumps utilize drop counters to control the injection rates. The drop counter, a typical through-beam sensor, is composed of a light emitter and a receiver facing each other, with the drip chamber situated in between. By overlaying a light signal of the same wavelength as the authentic light source, the authors could remotely manipulate the injection rate of the victim's infusion pump.

Similarly, Shin et al. [34] showed that multiple models of wide-area smoke detectors, which are retro-reflective sensors typically installed within critical infrastructures, could be remotely deceived to induce false alarms and suppress real alarms by injecting malicious optical signals into the victim sensors. Both saturation and spoofing attack techniques were utilized in these studies to manipulate victim sensors. Despite the critical nature of these attacks, none of the targeted products was equipped with anti-saturation/spoofing measures. Particularly for spoofing attacks, no defensive scheme had been previously suggested that could differentiate between malicious optical signals and authentic signals, which highlights the novelty of our work.

Attacks against Other Types of Active Sensors Several existing works have demonstrated sensor attacks against lidars, classified as diffuse-reflective sensors. Petit et al. [30] presented sensor attacks against an automotive lidar model equipped with a vision sensor. They concluded that it was possible to induce fake points within the lidar output by receiving authentic light pulses from the victim lidar and transmitting back malicious pulses simulating fake echoes. This study was later extended by Shin et al. [33] and Cao et al. [11], who demonstrated the feasibility of inducing fake points closer to the victim lidar than the spoofing device and exploiting the backend system processing the raw point cloud, respectively. Currently, Lightbox cannot defend against sensor attacks on lidars (a type of diffuse-reflective sensor), because reflections by arbitrary objects could alter the spectrum of the light beam, and the sensing distance is not fixed. This is an important consideration for future work, especially given the security concerns of self-driving cars using lidars.

Sensor attacks are not limited solely to the optical domain. Following Petit et al., Yan et al. [52] conducted research on sensors mounted on a Tesla Model S, including cameras, ultrasonic sensors, and radars. Moreover, plausible attacks against ultrasonic sensors were further investigated by Xu et al. [51]. Shoukry et al. [36] demonstrated that the Anti-lock Braking System (ABS) could be deceived via sensor spoofing by injecting arbitrary values into the wheel-speed sensor. Although Lightbox is only applicable to optical sensors, a similar approach might be effective for sensors utilizing other types of physical signals, as long as an analog for fingerprinting can be identified. This can serve as a direction for future research.

Other Methods of Defense against Sensor Attacks Based on Signal Fingerprinting Shoukry et al. proposed PyCRA, whose purpose is similar to that of Lightbox. PyCRA can detect sensor spoofing attacks on general active sensors by adding sudden notches to the emitted signal at random instances. The underlying principle is that the attacker would be incapable of reacting instantaneously due to the inherent delay involved [36]. However, Shin et al. pointed out that PyCRA implementations on analog-digital sensors could be circumvented unless the sensor utilized impractically high sampling rates [35]. This makes PyCRA's adoption in real-world systems challenging due to the scarcity of purely analog systems. Unlike PyCRA, Lightbox does not depend on temporal aspects of received signals and thus does not require impractically high processing speeds.

Outside the domain of sensor attacks, many fingerprinting-based studies have been conducted for device authentication. Cheng et al. validated the potential for distinguishing between CPU chips based on magnetic induction sensors [13]. Though not proposed as a defense technique against sensor attacks, Toh et al. suggested that light sources could be fingerprinted using the spectra of light signals [44]. Compared to our work, Toh et al.'s work is considerably limited, because it uses a multi-class classifier only in through-beam mode with 10 LEDs. Zhuang et al. demonstrated that a fake base station attack could be detected based on the fingerprinting of radio-frequency signals [56]. Further studies on fingerprinting have been suggested to classify mobile devices based on the unique features of in-built hardware components, such as microphones and speakers [3, 14, 54, 57].

9 CONCLUDING REMARKS

In this article, we propose Lightbox, an effective method for detecting sensor attacks on photoelectric sensors, based on fingerprinting the spectrum of received light signals. We demonstrate that Lightbox can successfully detect (1) saturation attacks, (2) simple spoofing attacks, and (3) sophisticated attacks. The real-world application of Lightbox is evident in safety-critical photoelectric sensor applications. For instance, a manufacturer could deploy Lightbox to detect sensor attacks, aiding in the prevention of accidents. Despite potential challenges, such as the need for sufficient reference data for model training and the necessity for Lightbox to operate effectively in an embedded system, we propose solutions. Moreover, we show an end-to-end scenario of Lightbox in real-world applications using Jetson TX2 and find that data near the spectrum peak functions as the most critical feature.

REFERENCES

[1] Amazon. [n. d.]. Optex Outdoor Dual Beam Photoelectric Detector. Retrieved from https://www.amazon.com/Optex-Outdoor-Dual-Photoelectric-Detector/dp/B007HKHJ8C

[2] Amazon. [n. d.]. Smoke Detector Fire Alarm Product. Retrieved from https://www.amazon.com/Detector-Ardwolf-Photoelectric-Battery-Powered-Included/dp/B071DQXW3W

[3] Gianmarco Baldini, Gary Steri, Franc Dimc, Raimondo Giuliani, and Roman Kamnik. 2016. Experimental identification of smartphones using fingerprints of built-in micro-electro mechanical systems (MEMS). *Sensors* 16, 6 (2016).

[4] BANNER. [n. d.]. Basics of Photoelectric Sensing. Retrieved from https://stevenengineering.com/tech_support/PDFs/04PHREF.pdf

[5] BANNER. [n. d.]. Liquid Leak Detection with a QS18. Retrieved from https://www.bannerengineering.com/my/en/solutions/other/liquid-leak-detection-with-a-qs18.html

[6] BANNER. [n. d.]. Q4X Series Sensor. Retrieved from https://www.bannerengineering.com/us/en/products/sensors/photoelectric-sensors/q4x-laser-distance-sensor.html

[7] BANNER. [n. d.]. Q5X Series Sensor. Retrieved from https://www.bannerengineering.com/us/en/products/sensors/photoelectric-sensors/q5x-series.html

[8] BANNER. [n. d.]. QS18 Series All Purpose Photoelectric Sensor. Retrieved from https://www.bannerengineering.com/my/en/solutions/other/conveyor-jam-detection.html

[9] BANNER. [n. d.]. WORLD-BEAM Q20 Series Sensor. Retrieved from https://www.bannerengineering.com/us/en/products/sensors/photoelectric-sensors/q20-series.html

[10] BANNER. [n. d.]. WORLD-BEAM QS30 Series Sensor. Retrieved from https://www.bannerengineering.com/us/en/products/sensors/photoelectric-sensors/qs30-series.html

[11] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z. Morley Mao. 2019. Adversarial sensor attack on LiDAR-based perception in autonomous driving. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS'19)*. Association for Computing Machinery, New York, NY, 15 pages. DOI: https://doi.org/10.1145/3319535.3339815

[12] Ruchir Chauhan. 2014. *A Platform for False Data Injection in Frequency Modulated Continuous Wave Radar*. Ph. D. Dissertation. Utah State University.

[13] Yushi Cheng, Xiaoyu Ji, Juchuan Zhang, Wenyuan Xu, and Yi-Chao Chen. 2019. DemiCPU: Device fingerprinting with magnetic signals radiated by CPU. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*.

[14] Anupam Das, Nikita Borisov, and Matthew Caesar. 2014. Do you hear what I hear? Fingerprinting smart devices through embedded acoustic components. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*.

[15] ELECTRO SENSORS. [n. d.]. Single-beam Safety Sensor E3ZS/E3FS. Retrieved from https://www.fa.omron.com.cn/data_pdf/cat/e3zs_e3fs_ds_e_10_4_csm1260.pdf?id=1519

[16] Penn Engineering. [n. d.]. Generic PCA Infusion Pump Reference Implementation. Retrieved from https://rtg.cis.upenn.edu/medical/gpca/rtg-gpca-video-v1.html

[17] F. Ernst and P. Pirouz. 1988. Formation of planar defects in the epitaxial growth of GaP on Si substrate by metal organic chemical-vapor deposition. *J. Appl. Phys.* 64, 9 (1988).

[18] Blaise Gassend, Dwaine Clarke, Marten Van Dijk, and Srinivas Devadas. 2002. Silicon physical random functions. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*. ACM.

[19] Jun Gou, Qingchen Niu, Kai Liang, Jun Wang, and Yadong Jiang. 2018. Frequency modulation and absorption improvement of THz micro-bolometer with micro-bridge structure by spiral-type antennas. *Nanosc. Res. Lett.* 13, 1 (2018), 1–9.
[20] Intel. [n. d.]. Intel RealSense LiDAR Camera L515. Retrieved from https://www.intelrealsense.com/lidar-camera-l515/
[21] Markus G. Kuhn and Ross J. Anderson. 1998. Soft tempest: Hidden data transmission using electromagnetic emanations. In *Proceedings of the International Workshop on Information Hiding*. Springer, 124–142.
[22] LEDinside. 2015. Color Consistency—One of the Challenges for Solid-state Lighting. Retrieved from https://www.ledinside.com/knowledge/2015/8/color_consistency_one_of_the_challenges_for_solid_state_lighting
[23] Kevin McGrattan, Simo Hostikka, Randall McDermott, Jason Floyd, Craig Weinschenk, and Kristopher Overholt. 2013. Fire dynamics simulator user's guide. *NIST Spec. Public.* 1019, 6 (2013).
[24] MongoDB. [n. d.]. MongoDB Server. Retrieved from https://www.mongodb.com/
[25] OceanInsight. [n. d.]. STS-NIR-L-10-400-SMA. Retrieved from https://www.oceaninsight.com/products/spectrometers/microspectrometer/sts-series/sts-nir-l-10-400-sma/?qty=1
[26] Omron. [n. d.]. Overview of Photoelectric Sensors. Retrieved from https://www.ia.omron.com/support/guide/43/introduction.html
[27] Youngseok Park, Yunmok Son, Hocheol Shin, Dohyun Kim, and Yongdae Kim. 2016. This ain't your dose: Sensor spoofing attack on medical infusion pump. In *Proceedings of the 10th USENIX Workshop on Offensive Technologies (WOOT'16)*. USENIX Association. Retrieved from https://www.usenix.org/conference/woot16/workshop-program/presentation/park
[28] Pepperl Fuchs. [n. d.]. Retroreflective Sensor MLV12-54/47/92. Retrieved from https://www.pepperl-fuchs.com/korea/ko/classid_11.htm?view=productdetails&prodid=7224
[29] Pramuditha Perera and Vishal Patel. 2019. Learning deep features for one-class classification. *IEEE Transactions on Image Processing* 28, 11 (2019), 5450–5463.
[30] Jonathan Petit, Bas Stottelaar, Michael Feiri, and Frank Kargl. 2015. Remote attacks on automated vehicles sensors: Experiments on camera and LiDAR. *Black Hat Eur.* 11 (2015).
[31] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
[32] Michael Royer. 2014. Color Stability of LEDs: Understanding the Basics. Retrieved from https://www1.eere.energy.gov/buildings/publications/pdfs/ssl/royer_stability_lightfair2014.pdf
[33] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. 2017. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems*. Springer.
[34] Hocheol Shin, Juhwan Noh, Dohyun Kim, and Yongdae Kim. 2020. The system that cried wolf: Sensor security analysis of wide-area smoke detectors for critical infrastructure. *ACM Trans. Priv. Secur.* 23, 3, Article 15 (June 2020), 32 pages. DOI: https://doi.org/10.1145/3393926
[35] Hocheol Shin, Yunmok Son, Youngseok Park, Yujin Kwon, and Yongdae Kim. 2016. Sampling race: Bypassing timing-based analog active sensor spoofing detection on analog-digital systems. In *Proceedings of the 10th USENIX Workshop on Offensive Technologies (WOOT'16)*.
[36] Yasser Shoukry, Paul Martin, Yair Yona, Suhas Diggavi, and Mani Srivastava. 2015. PyCRA: Physical challenge-response authentication for active sensors under spoofing attacks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*.
[37] Yun Shuai, Yongzhi He, Nguyen T. Tran, and Frank G. Shi. 2010. Angular CCT uniformity of phosphor converted white LEDs: Effects of phosphor materials and packaging structures. *IEEE Photon. Technol. Lett.* 23, 3 (2010).
[38] Softnoze. [n. d.]. Photoelectric Sensors Theory of Operation. Retrieved from https://www.softnoze.com/downloads/Sensor%20Basics%204.pdf
[39] Ji-Su Son, Yoshio Honda, and Hiroshi Amano. 2014. Growth of low-defect-density nonpolar a-plane GaN on r-plane sapphire using pulse NH 3 interrupted etching. *Optics Expr.* 22, 3 (2014).
[40] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. Light commands: Laser-based audio injection attacks on voice-controllable systems. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security'20)*.
[41] The Great Soviet Encyclopedia. [n. d.]. Modulation of Light. Retrieved from https://encyclopedia2.thefreedictionary.com/Modulation+of+Light
[42] Thorlabs. [n. d.]. Compact CCD Spectrometers. Retrieved from https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=3482
[43] Thorlabs. [n. d.]. Cosine Corrector for SMA-Connectorized Fiber. Retrieved from https://www.thorlabs.com/thorproduct.cfm?partnumber=CCSA1

[44] A. Toh, L. Yang, K. Sakiyama, and T. Sugawara. 2019. Fingerprinting light emitting diodes using spectrometer. *Electron. Lett.* 55, 24 (2019).
[45] Nguyen The Tran, Jiun Pyng You, and Frank G. Shi. 2009. Effect of phosphor particle size on luminous efficacy of phosphor-converted white LED. *J. Lightw. Technol.* 27, 22 (2009).
[46] Ingrid Verbauwhede and Roel Maes. 2011. Physically unclonable functions: Manufacturing variability as an unclonable device identifier. In *Proceedings of the 21st Edition of the Great Lakes Symposium on Great Lakes Symposium on VLSI*. ACM.
[47] Vicardio. [n. d.]. The Principle of Blood Pressure Detector. Retrieved from http://www.vicardio.care/how-it-works/
[48] P. Visconti, D. Huang, M. A. Reshchikov, F. Yun, R. Cingolani, D. J. Smith, J. Jasinski, W. Swider, Z. Liliental-Weber, and H. Morkoc. 2002. Investigation of defects and surface polarity in GaN using hot wet etching together with microscopy and diffraction techniques. *Mater. Sci. Eng.: B* 93, 1-3 (2002).
[49] Waymo. [n. d.]. Google Waymo. Retrieved from https://waymo.com/waymo-driver/
[50] D. S. Wuu, W. K. Wang, K. S. Wen, S. C. Huang, S. H. Lin, S. Y. Huang, C. F. Lin, and R. H. Horng. 2006. Defect reduction and efficiency improvement of near-ultraviolet emitters via laterally overgrown GaN on a GaN/patterned sapphire template. *Appl. Phys. Lett.* 89, 16 (2006).
[51] W. Xu, C. Yan, W. Jia, X. Ji, and J. Liu. 2018. Analyzing and enhancing the security of ultrasonic sensors for autonomous vehicles. *IEEE Internet Things J.* 5, 6 (Dec 2018). DOI: https://doi.org/10.1109/JIOT.2018.2867917
[52] Chen Yan, Wenyuan Xu, and Jianhao Liu. 2016. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle. *DEF CON* 24 (2016).
[53] Huai Zheng, Xiaobing Luo, Run Hu, Bin Cao, Xing Fu, Yiman Wang, and Sheng Liu. 2012. Conformal phosphor coating using capillary microchannel for controlling color deviation of phosphor-converted white light-emitting diodes. *Optics Expr.* 20, 5 (2012).
[54] Zhe Zhou, Wenrui Diao, Xiangyu Liu, and Kehuan Zhang. 2014. Acoustic fingerprinting revisited: Generate stable device id stealthily with inaudible sound. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*.
[55] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* 23, 4 (1997), 550–560.
[56] Zhou Zhuang, Xiaoyu Ji, Taimin Zhang, Juchuan Zhang, Wenyuan Xu, Zhenhua Li, and Yunhao Liu. 2018. FBSleuth: Fake base station forensics via radio frequency fingerprinting. In *Proceedings of the Asia Conference on Computer and Communications Security*.
[57] Ling Zou, Qianhua He, and Junfeng Wu. 2017. Source cell phone verification from speech recordings using sparse representation. *Digit. Sig. Process.* 62 (2017).

Received 10 January 2023; revised 18 July 2023; accepted 31 July 2023