LiDAR Spoofing Meets the New-Gen: Capability Improvements, Broken Assumptions, and New Attack Strategies

Takami Sato*†, Yuki Hayakawa*‡, Ryo Suzuki*‡, Yohsuke Shiiki*‡, Kentaro Yoshioka‡, Qi Alfred Chen†

†University of California, Irvine, Department of Computer Science
‡Keio University, Department of Electronics and Electrical Engineering
†{takamis, alfchen}@uci.edu, ‡{hykwyuk, suzuki.ryo, kyoshioka47}@keio.jp, ‡shiiki@iskr.elec.keio.ac.jp

Abstract—LiDAR (Light Detection And Ranging) is an indispensable sensor for precise long- and wide-range 3D sensing, which directly benefited the recent rapid deployment of autonomous driving (AD). Meanwhile, such a safety-critical application strongly motivates its security research. A recent line of research finds that one can manipulate the LiDAR point cloud and fool object detectors by firing malicious lasers against LiDAR. However, these efforts face 3 critical research gaps: (1) considering only one specific LiDAR (VLP-16); (2) assuming unvalidated attack capabilities; and (3) evaluating object detectors with limited spoofing capability modeling and setup diversity.

To fill these critical research gaps, we conduct the first large-scale measurement study on LiDAR spoofing attack capabilities on object detectors with 9 popular LiDARs, covering both first- and new-generation LiDARs, and 3 major types of object detectors trained on 5 different datasets. To facilitate the measurements, we (1) identify spoofer improvements that significantly improve the latest spoofing capability, (2) identify a new object removal attack that overcomes the applicability limitation of the latest method to new-generation LiDARs, and (3) perform novel mathematical modeling for both object injection and removal attacks based on our measurement results. Through this study, we are able to uncover a total of 15 novel findings, including not only completely new ones due to the measurement angle novelty, but also many that can directly challenge the latest understandings in this problem space. We also discuss defenses.

I. Introduction

LiDAR (Light Detection And Ranging) is one of the most innovative sensors in the past decade. By shooting a laser pulse and measuring its reflection, LiDAR can provide a detailed 3D understanding of the surrounding environment. Autonomous Driving (AD) is one of the most benefited applications of the high-speed and high-precision sensing of LiDARs. After LiDAR showed its effectiveness in the 2007 DARPA Urban Challenge [1], it has been widely recognized as an essential sensor for Level-4 AD and has been adopted in almost all recent robotaxi services (Waymo One [2], Cruise [3]) and AD vehicles operating in the US [4, 5]. While highly beneficial to our everyday life and society, AD is also highly security-critical as even a small operational error can cause fatal
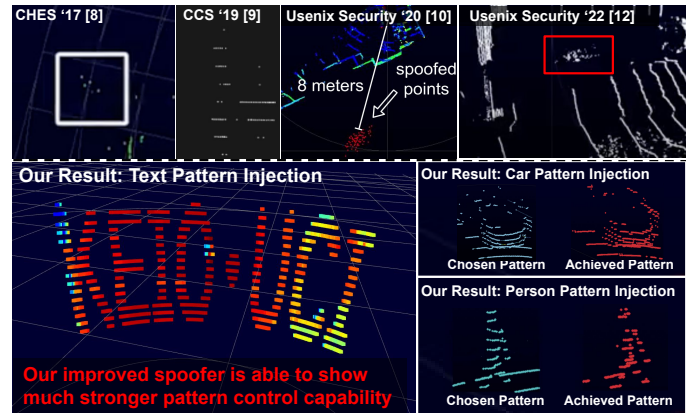
*co-first authors

Figure 1: Demonstration of the Chosen Pattern Injection (CPI) attack capability with >6,000 spoofed points with our improved LiDAR spoofer. This significantly improves the spoofing attack capability from prior works: [8] (∼10 points), [9] (∼20 points) [10] (∼200 points), and [12] (∼200 points).

consequences [6]. To address this, numerous researchers have been conducting security analyses on LiDARs [7]–[13] due to their critical role in AD perception. The major security concern of LiDARs is the fundamental vulnerability against malicious laser shooting, or *LiDAR spoofing attacks*. The recent research along this line [10, 12, 14] found that such attacks can cause both false positives (injecting a non-existing fake object) and false negatives (removing an existing object). However, we find that there are 3 critical research gaps in these prior efforts:

Considering only one specific LiDAR: Velodyne VLP-16 [15] has been dominantly used in the prior works since it is viewed as a *de facto* choice for LiDAR spoofing evaluation after the first practical spoofing attack was proposed in 2017 [8]. The following works thus all evaluate their attacks only on VLP-16 [9, 10, 12, 13] or use the attack capability on VLP-16 to justify the validity of their threat model [14, 16]. Although these results are valid on VLP-16, there is no guarantee that these results are still valid in more recent LiDARs, known as next-generation or new-generation (or *new-gen*) LiDARs [17], as opposed to the first-generation (or *first-gen*) ones such as VLP-16. The new-gen LiDARs have more advanced security-related features, such as laser timing randomization and pulse fingerprinting. Prior works [8, 9, 13] partially discussed some of them as potential defenses, but none of them experimentally studied their impact and effectiveness against LiDAR spoofing.

Assuming unvalidated attack capabilities: So far, all prior works on fake object injection [9, 10, 12] assume a *Chosen Pattern Injection (CPI)* attack capability, i.e., an attacker can spoof a specifically chosen point cloud pattern that was carefully optimized/identified offline beforehand. For example, the Adv-LiDAR attack [9] is a white-box adversarial attack that needs a specific pattern to trigger a vulnerability in deep neural networks (DNNs). The black-box attacks in [10, 12] also require a specific pattern in the shape of a vehicle surface. However, none of them have systematically studied how achievable this is in practice. The top area of Fig. 1 shows the demonstrated spoofing capabilities so far from prior works [8]–[10, 12]. As shown, none of them were able to show strong pattern control capabilities, not to mention to provide a systematic quantification of such a capability to ensure valid security analysis on the object detector side.

Evaluating object detectors with limited spoofing capability modeling and setup diversity: To study the spoofing attack impacts on the object detector side, prior works typically leverage a mathematical modeling of the spoofing capability to achieve high analysis scalability and flexibility [9, 10, 12]. However, the modelings used so far not only fail to correctly capture real-world attack characteristics such as spoofing inaccuracies (detailed in §IV-A2, which can significantly bias the analysis results as shown in §IV-B3), but also lack the coverage of more recent LiDAR features that can significantly change the spoofing capabilities (§IV-A4). Moreover, the object detector model setup is also limited, e.g., no prior works have evaluated the same model design trained on different training datasets, which we found can significantly change the model robustness analysis results (§IV-B3 and §V-B2).

To fill these critical research gaps, we conduct the first large-scale measurement study on LiDAR spoofing attack capabilities against object detection. Specifically, our study is driven by 3 key novel research questions (RQ):
RQ1: *Are the common design-level assumptions made in prior LiDAR spoofing attacks actually realizable? If so, can they also hold for the more recent new-gen LiDARs?*
RQ2: *Do different types of LiDARs, especially the new-gen ones with security-related features, have different vulnerability characteristics to LiDAR spoofing attacks?*
RQ3: *Does the vulnerability status of popular object detectors to spoofing attacks significantly change due to the new-gen LiDAR features and different model setups?*

To comprehensively address these RQs, we devote significant engineering efforts to reproduce and/or improve the current state-of-the-art attacks: For RQ1, to properly explore the potential to achieve the CPI attack capability, we improve the spoofing device on their optical and electronics parts (§III-D), which enables us to be the first to demonstrate and quantify the CPI attack capability, which is commonly assumed but never clearly demonstrated in prior works [9, 10, 12]–[14] and shown in Fig. 1. For RQ2, to adequately measure the vulnerability status of new-gen LiDARs, we explore *asynchronized* (§II-C) spoofing attacks since synchronized ones are directly foiled by their common security-related features such as timing randomization and pulse fingerprinting (§IV-A3). However, since all existing works only consider first-gen LiDARs, their designs predominantly focus on synchronized attacks [8]–[10, 12, 13], leaving the asynchronized attack design space

under-explored. For RQ3, to enable large-scale evaluation of LiDAR spoofing attack capabilities against different object detectors, we perform novel mathematical modeling of the spoofing attack capabilities on different types of LiDARs based on our measurements for RQ1 and RQ2, which is not only the first to perform such modelling for first- and new-gen LiDARs, but also the first to model for object removal attacks.

In the measurement study, we cover major types of LiDAR spoofing attacks against (1) 9 popular LiDAR models in total, covering both the classic first-gen ones (e.g., VLP-16) and the new-gen ones with new security-related features; and (2) 3 major types of object detectors trained on 5 different datasets. To the best of our knowledge, this is the first large-scale measurement study on this topic in terms of *LiDAR numbers* (none of the prior works studied over 1 LiDAR model, while we study 9), *LiDAR types* (all prior works concentrate on first-gen ones, and we are the first to study the new-gen ones), and also *training datasets* (no prior works studied the same model design with over 1 dataset, while we study 5).

Through this study, we are able to identify a total of *15 novel findings*, which include not only completely new ones due to the measurement angle novelty (e.g., measurements on new-gen LiDARs), but also many that directly challenge the latest understandings in this problem space. For example:

- We find that with more careful spoofer optics and electronics implementations, an attacker actually does not really need to exploit object detector-level vulnerabilities to achieve a near-front road object injection, which is a common assumption made in prior works [9, 10, 12];
- We find that VLP-16 is actually the *only* LiDAR model for which the CPI attack capability assumption is feasible, which is another key design assumption made in latest prior works [9, 10, 12];
- We find that the new-gen LiDAR features previously expected to be capable of largely mitigating or even preventing spoofing (e.g., timing randomization [9, 13], pulse fingerprinting [8]) may not be effective as expected (§IV-A4, §IV-B3);
- We find that the latest synchronized object removal attack can no longer be applied to the new-gen LiDARs, but there exist asynchronized object removal attacks that can overcome such a limitation and can lead to a similar level of practical attack capabilities (§V).

In summary, our study has the following contributions:

- We conduct the first large-scale measurement study on LiDAR spoofing attack capabilities on object detectors with 9 LiDARs, covering both first- and new-generation ones, and 3 major types of object detectors. We are also the first to investigate the impacts of new security-related features in more recent LiDARs from the security perspective.
- To facilitate the measurements, we not only identify spoofer improvements that significantly improve the latest spoofing capability, but also identify a new asynchronized object removal attack that overcomes the applicability limitation of the latest method to new-gen LiDARs while having a similar level of practical attack capability.
- To facilitate the object detector-level measurements, we perform novel mathematical modeling of the spoofing capabilities for both object injection and removal attacks

based on our measurement results, which is the first to model for both first- and new-gen LiDARs and also to model for object removal attacks.

- Our study is able to uncover a total of 15 novel findings, including not only completely new ones due to the measurement angle novelty, but also many that can directly challenge the latest understandings in this problem space.

Data/code release. All data, code, and hardware designs (e.g., for the new spoofer) are released at **https://sites.google.com/view/cav-sec/new-gen-lidar-sec**.

II. BACKGROUND AND RELATED WORKS

A. LiDAR Basics and Recent Trend

LiDAR is an active sensor that can obtain high-resolution 3D point cloud data of the surrounding environment. The most common type of LiDAR today is direct time-of-flight (dToF) LiDAR, which fires laser pulses to the environment and uses their reflections to actively measure (or "scan") the surrounding objects. Specifically, for each laser pulse, it uses the time difference between the firing and reflection to obtain the 3D position of a "point" on the object surface. By performing laser firing to different horizontal angles (*azimuth*) and vertical angles (*altitude*), the point measurements thus form a "point cloud" (Fig. 1). Such LiDARs typically use laser peak power (e.g., ∼10 W) significantly higher than sunlight even after a long flight, which allows them to detect objects more than 200 meters away even under high ambient lights.

Recent Trend: New-Gen LiDARs. Early Velodyne LiDARs such as VLP-16 [15] and VLP-32c [18] are commonly known as the first-generation (*first-gen*) LiDARs [17], which naively integrate the classic point-wise laser ranging systems. The advent of first-gen LiDARs has significantly boosted critical real-world applications such as autonomous driving (AD), but its complex mechanical design increases costs and limits scalability. To overcome the limitations, the new-generation (*new-gen*) LiDARs [17] mount all components, such as the photodetector and the readout circuitry, on a single chip (called a system-on-chip (SoC) approach). This not only reduces the cost and improves the scalability of the system, but also allows new designs of laser firing and receiving. For example, microelectromechanical systems (MEMS) LiDAR [19] move a mirror to scan a wide azimuth range instead of mechanical rotation. Flash LiDARs [20, 21] fire a broad laser that covers the entire field of view (FOV) and calculates the distance by compensating its weak return laser by accumulation. Moreover, the SoC approach allows more complex signal processing, such as a large number of simultaneous laser firings [22], laser timing randomization [22, 23], and fingerprinting [24], to be robust against challenging environments, e.g., multiple LiDARs operating adjacent to each other.

All in all, the recent new-gen LiDARs substantially improved the electronics and the scanning mechanisms even though they still have the same basic design: laser firing and receiving. However, none of the prior works on LiDAR spoofing attacks have studied the security property of such new-gen LiDARs; *all* of them are performed on only the first-gen rotational ones, with a predominate focus on in fact *only 1 specific LiDAR model*: VLP-16 [8]–[10, 12, 13]. The major design differences between first- and new-gen LiDARs

Table I. Taxonomy of existing LiDAR spoofing attacks against 3D object detection. Sync. and Async. refer to synchronized and asynchronized spoofing techniques (§II-C).

Spoofing Technique	Object Detection Model-Level Attack Effect	
	Object Injection	Object Removal
Sync. (White-box)	Adv-LiDAR [9], Occlusion [10], Frustum [12]	PRA [13], ORA [7]
Async. (Black-box)	Relay* [7]	Saturating† [8], HFR (§III-C)

* Cannot inject fake objects (1) closer than the spoofer itself, and (2) in a different angle towards the victim than the spoofer itself.
† Only show removal of a 41×42 cm^2 metal plate in short duration (4 sec)

are likely to cause significant differences in their security characteristics, which thus motivates this study.

B. Object Detection on Point Cloud

In real-world applications such as autonomous driving (AD), object detection is no doubt one of the most critical roles for LiDAR. As in other computer vision areas, 3D object detection on point cloud is substantially benefited by the recent progress of DNN. However, traditional DNN architectures such as CNN cannot be directly applied due to the irregular and unordered structure of point clouds [25]. To handle such data complexity, 3 major types of 3D object detection methods are widely adopted: *voxel-based*, *point-based*, and *point voxel-based* methods [26]. More details are in Appendix D.

C. LiDAR Spoofing Attacks against Object Detection

Due to the basic sensing mechanism of laser firing and receiving, LiDARs are fundamentally vulnerable to malicious laser shooting, or *LiDAR spoofing attacks*. Specifically, such attacks work by using an external attack device ("spoofer") to fire laser pulses back to the victim LiDAR in order to manipulate the time measurements of the laser-receiving events and thus the corresponding 3D position measurements (§II-A). Such point position manipulations can thus be strategically used to achieve attack effects on the 3D object detector side, more specifically (1) *Object injection*, e.g., by moving the positions of a set of points to form a 3D pattern to cause a false-positive detection of a non-existing object; and (2) *Object removal*, e.g., by moving the points originally on an object to elsewhere to cause false-negative detection of such an object.

1) Attack Taxonomy: Table I shows a taxonomy of existing LiDAR spoofing attacks on object detection based on the LiDAR spoofing technique (specifically, the requirement of *synchronization*) and the targeted object detector-level attack effect. *Synchronization* means to synchronize the malicious laser firing timing with the victim LiDAR scanning (i.e., laser firing) timing, which can enable precise control of the attack laser-receiving timing and thus the corresponding positions of the spoofed points. Fig. 2 illustrates the synchronized and asynchronized spoofing techniques with the latest attack capabilities. As shown, to achieve synchronization, the attacker needs to use an extra device (photodetector, or PD, in Fig. 2) to first learn the current state of the victim LiDAR scanning in the real time, and then use the victim LiDAR's scanning pattern to derive future victim laser-firing timings for synchronized attack laser-firing. More detailed explanations are in Appendix A. This process requires precise and predictable knowledge of the victim LiDAR's scanning pattern beforehand, which thus can be viewed as a *white-box* LiDAR attack. Those that do not
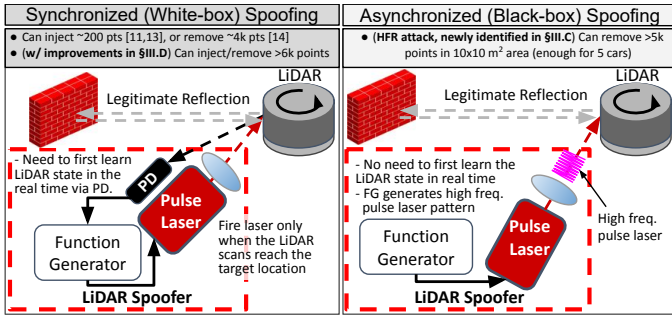
Synchronized (White-box) Spoofing	Asynchronized (Black-box) Spoofing
• Can inject ~200 pts [11,13], or remove ~4k pts [14] • (w/ improvements in §III.D) Can inject/remove >6k points	• (HFR attack, newly identified in §III.C) Can remove >5k points in 10x10 m² area (enough for 5 cars)

Figure 2: Illustration of the synchronized and asynchronized LiDAR spoofing techniques with the latest attack capabilities. Synchronized spoofing needs white-box knowledge of the victim LiDAR scanning patterns and an extra device (Photodetector, or PD) for synchronization (§II-C), while asynchronized spoofing does not need these (i.e., black-box LiDAR attack).

assume synchronization (i.e., *asynchronized* attacks) can be applied without knowing anything about the LiDAR internal scanning logic, which can be viewed as *black-box* attacks.

Spoofing Attacks for Object Injection. The very first LiDAR spoofing attack was discovered in 2015 [7], which found that shooting lasers to a victim LiDAR can inject ≥200 points and achieve fake object injection at the object detector side. This attack works by relaying the laser received from the victim LiDAR, which thus does not require synchronization. However, this also makes it impossible to inject objects (1) in closer positions than the spoofer itself, and (2) in a different angle towards the victim than the spoofer. These limitations make it not very useful in real-world attack scenarios since at the time of spoofing, the attackers would hope that the most threatening object to the victim vehicle is the induced fake objects instead of the attacker herself [8].

To address such limitations, synchronized spoofing [8] is proposed to use the *synchronization* process described above and also illustrated in Fig. 2. The early-stage attack designs can inject only 10 spoofed points [8], but the following works progressively increase the number of spoofed points to 60 [9] and 200 [10], and show that the 60 and 200 spoofed points are enough to cause fake object injection on latest object detectors. After the success, the attack capability of injecting 200 points under the Chosen-Pattern Injection (CPI) assumption (detailed later in §II-C2) becomes the *de facto* threat model in the following works [12, 14, 16]. However, none of them have (1) demonstrated and quantified the CPI attack capability, and (2) studied their attack effect on new-gen LiDARs, which have new features such as laser timing randomization and pulse fingerprinting that can directly challenge their basic design assumption — the requirement of synchronization. In this paper, we fill both of these critical research gaps.

Spoofing Attacks for Object Removal. With the success to achieve object injection, more recent works start to explore using LiDAR spoofing for object removal. Specifically, saturating attack [8] is the first to show the object removal effect. In this attack, instead of firing pulsed lasers like the prior works above, it fires a strong *continuous* laser to the victim LiDAR to indirectly cause measurement errors of laser-receiving events. However, due to the requirement of maintaining continuous high-power laser, it is physically difficult for the attack laser beam to (1) achieve a large receiver area coverage at the victim

LiDAR side with high intensity, and (2) maintain the attack effect for a long time. For example, the demonstrated object removal effect is only about removing a 41×42 cm² metal plate, lasting <4 seconds, which thus makes it highly limited in real-world attack scenarios such as when attacking AD. For example, the saturating attack (0.8 W average power) needs roughly 10 times higher average power than the HFR attack (80 W peak power) (§III-C) to achieve similar attack effectiveness, but such a powerful diode is not publically available so far.

To address such limitations, similar to the object injection side, more recent works start to leverage synchronized spoofing techniques. Specifically, the ORA attack (object removal attack) [14] found that using the synchronized spoofing capability of injecting 200 points under the CPI assumption as mentioned above, attackers can fool 3D object detectors by strategically injecting spoofed points inside the target object's bounding box, since the point cloud of legitimate objects should have points mostly on the object surface instead of inside. However, such removal effect still depends on the object detector-side vulnerabilities. Most recently, the physical removal attack (PRA) [13] found that the object removal effect can be more generally achieved by moving all points on a victim object to within the minimum operational threshold (MOT) of the victim LiDAR, which is common filtering mechanism to automatically discard points below a certain distance. This filtering is implemented in most LiDARs because the LiDAR measurement at a very close distance is generally inaccurate since the ToF can be too short to distinguish from errors due to inadequate calibration. This work shows the capability of removing ~4k points and thus can remove critical road objects such as pedestrians, which is thus the current state-of-the-art in object removal attack capability.

However, similar to the object injection attack side, none of these prior works have considered the new-gen LiDARs that may come with features that can break some of their fundamental design assumptions. In this paper, we find that (1) PRA can no longer be applied to new-gen LiDARs due to the requirement of synchronization (§V-A), and (2) the assumption of non-zero MOT for PRA may not necessarily hold (e.g., for XT32 [24], §IV-A5). We further identify a new object removal attack adapted from the saturation attack, which does not require synchronization but can still achieve practical object removal effects (can remove >5k points in 10×10 m² area, which is enough for 5 cars as shown later in §V-B2).

2) Chosen-Pattern Injection (CPI) Attack Capability: So far, all prior works using synchronized spoofing for object injection explicitly or implicitly assume that at the actual attack time the attacker is able to accurately inject a specific spoofed point cloud pattern chosen by the attacker beforehand, e.g., via various kinds of off-line optimization/identification processes [9, 10, 12]. In this paper, we thus explicitly define this as the *Chosen Pattern Injection (CPI)* attack capability.

Such an attack capability is theoretically achievable since the synchronized LiDAR spoofing technique should by design be capable of precisely controlling the position of each spoofed point. However, so far no prior works have systematically studied how achievable this is in practice, not to mention to provide a systematic quantification of the pattern control capability, which is necessary for achieving valid security analysis on the object detector side. In this paper, we thus

perform the first measurement study to fill this critical gap, with the coverage of both first- and new-gen LiDARs.

D. Threat Model

We follow the same threat model as in prior works [9, 10, 12], i.e., the attacker fires malicious lasers from their spoofer to the victim LiDAR (§II-C, Fig. 2). As described in [12], the spoofer device can be at a front vehicle, vehicle in the next lane, or a roadside in AD scenarios.

III. MEASUREMENT STUDY SETUP

A. Targeted LiDARs

Table II summarizes the LiDARs involved in this measurement study. As shown, we are able to cover 9 popular LiDARs in total, including 3 first-gen ones, 6 new-gen ones, an operating range from 9 m to 300 m, 3 scanning types (Rotating, MEMS, and Flash), and 3 security-related features (simultaneous laser firing, timing randomization, and fingerprinting). This makes this work the largest-scale measurement study on this topic so far in terms of both LiDAR numbers (none of the prior works studied over 1 LiDAR model, while we study 9) and LiDAR types (all prior works concentrate on first-gen ones, and we are the first to study the new-gen ones).

B. Targeted Object Detection Models and Datasets

Our study includes 5 popular DNN-based 3D object detectors in total covering all the 3 major model types (§II-B): voxel-based (PointPillars [27], SECOND [28], and Part-A^2 [29]), point-based (3DSSD [30]), and point voxel-based (PV-RCNN [31]). In order to study the impact of the training dataset, we also prepare 5 different versions of PointPillars [27] trained on 5 different datasets (KITTI [32], Waymo [33], nuScenes [34], Lyft [35] datasets), and the dataset used in Apollo 6.0 [36]. This thus makes our measurement also the largest in terms of the coverage of object detection model architectures and training datasets. We directly use pre-training ones from MMdetection3D [37], OpenPCDet [38], or the Apollo repository [36].

C. Targeted LiDAR Spoofing Attacks

Our study targets spoofing attacks with the latest attack capabilities on both the object injection and removal sides (§II-C1). Specifically, for object injection, we reproduce the latest synchronized spoofing technique [9, 10, 13], and for object removal, we reproduce the latest physical removal attack (PRA) [13]. In addition, since the latest removal attack cannot be applied when synchronization is not possible (which we found to be the case for new-gen LiDARs in general, detailed in §IV-A3), we further identify and include in our measurement a new asynchronized spoofing technique for object removal, which is adapted from the saturation attack (§II-C1) but with much more practical object removal capabilities:

High-Frequency Removal (HFR): New Asynchronized Spoofing Technique for Object Removal. As described in §II-C1, due to the requirement of maintaining continuous high-power laser, the saturation attack is fundamentally limited in the removal area size and duration (e.g., 41×42 cm^2 in <4 seconds [8]), making it highly limited in real-world attack scenarios. Later works address this by using synchronization, but this also makes them inapplicable to new-gen LiDARs

(§IV-A3). To address this, we identify a new adaptation of the saturation attack, which still does not require synchronization, but instead of using high-power continuous laser, it uses *high-frequency pulsed laser*. This new attack is thus called *high-frequency removal (HFR) attack*. The key idea is to fire a large number of attack laser pulses to the victim LiDAR at a frequency that is higher than the laser-firing frequency of the victim LiDAR. This allows the attack laser to hit every laser-firing event of the victim LiDAR in the scanning range hit by the spoofer, which can thus achieve the spoofing effect for every points in that range without any knowledge of the victim scanning pattern (i.e., without requiring synchronization). Fig. 2 illustrates the spoofing mechanism, and Fig. 15 in Appendix illustrates the difference to the saturation attack. Due to the lack of synchronization, the receiving timing of the attack laser is random, and thus the spoofing effect will be moving each legitimate surface point of the targeted objects to a random position or undetectable area of the victim LiDAR (e.g., within MOT). As shown later in §V-B2, this can completely destruct the point cloud patterns at the original object position and thus cause the object removal effects.

D. LiDAR Spoofer Setup

Fig. 3 shows an overview of the LiDAR spoofer, its optics design, and the setups of the indoor and outdoor experiments to support our measurement study of the targeted spoofing attacks (§III-C). We generally follow the setup adopted in the latest works [9, 10, 12, 13], so we leave the details to Appendix B.

Spoofer Improvements. When reproducing the prior works, we also implemented several improvements on top of the latest spoofer setup to more properly study the spoofing attack capabilities. For instance, we find that inadequate optical design can significantly affect the number and angle coverage of spoofable points due to undesired diffusion and convergence of the attack laser beam. Specifically, if the laser beam is expanding, the intensity rapidly decays with distance, which thus limits the number of spoofable points; if the laser beam is converging, the diameter of the beam decreases with the distance, which thus makes it difficult to aim and cover a wide receiver angle. Ideally, the laser beam should be collimated without diffusion and convergence to achieve the target LiDAR with a minimum loss. To achieve this, we use a 25.4 mm focal length plano-convex lens with 1 inch (25.4 mm) diameter to cover the laser emitted by SPL90_3 [43], which has a maximum beam divergence angle of 25°. Since the diameter of the laser at the focal point is $\tan 25° \times 25.4 \times 2 = 23.7$ mm, we can cover it with the 1 inch (25.4 mm) lens. To precisely calibrate the lens setup, we develop a device that can use a hollow screw to adjust the distance between the laser diode (LD in Fig. 3) and the lens. Technically, this allows the spoofer to maintain a stably large number and angle coverage of spoofable points from hundreds of meters away.

Besides the optics side, we also notice that inadequate electronic implementations in the spoofer setup can introduce inaccurate laser detection timing (from LiDAR to PD in Fig. 3) and also long and unstable delays in the signal processing of the spoofer devices, which can thus significantly affect the attack capability in precisely controlling the placement of a spoofed point at a chosen 3D position. To address this, we improve the amplifier for the PD to increase the laser detection

Table II. The 9 LiDARs involved in our measurement study. 1st-G and New-G: First- and new-generation. FOV: Field of View.

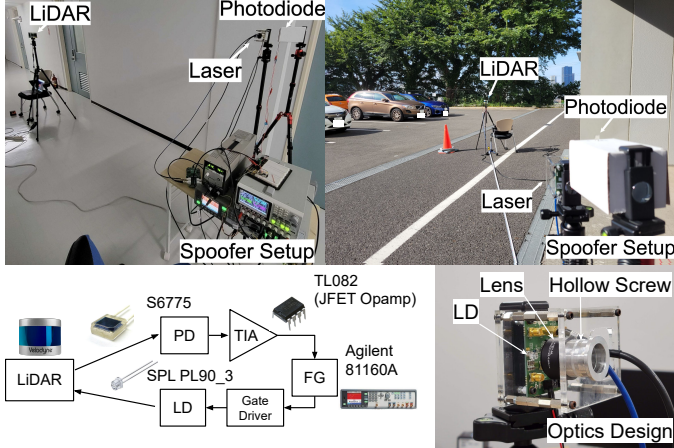|	Velodyne			Leddar	Ouster	Intel	Livox	Hesai	Robosense
	VLP-16 [15]	VLP-32c [18]	VLS-128 [39]	Pixell [40]	OS1-32 [22]	Realsense L515 [41]	Horizon [42]	XT32 [24]	Helios 5515 [23]
General Specs									
Gen. (year)	1st-G (2016)	1st-G (2017)	1st-G (2017)	New-G (2019)	New-G (2019)	New-G (2019)	New-G (2020)	New-G (2020)	New-G (2021)
Scanning Type	Rotating	Rotating	Rotating	Flash	Rotating	MEMS	MEMS	Rotating	Rotating
Wavelength	905 nm	905 nm	905 nm	905 nm	865 nm	860 nm	905 nm	905 nm	905 nm
Vertical FOV	30°	40°	40°	16°	45°	55°	25.1°	31°	70°
Horizontal FOV	360°	360°	360°	180°	360°	70°	81.7°	360°	360°
Max. Range [m]	100	200	300	56	120	9	260	120	150
Min. Range [m]	1	1	0.5	0.1	0.3	0.25	0.5	0	0.2
Vertical Channel	16	32	128	8	32	-	-	32	32
Security									
Simul. Firing	1	2	8	3	32	1	1	1	1
Timing Random.				✓	✓	✓	✓		✓
Fingerprinting								✓	

Figure 3: Overview of our LiDAR spoofer setup, the optics design, and the setup of the indoor and outdoor experiments. PD: Photodetector. TIA: Transimpedance amplifier. FG: Function generator. LD: Laser diode. More details in Appendix B.

accuracy, and also improve the function generator (FG) setup to allow more precise nanosecond-level configuration and calibration. As shown in Fig. 1 and detailed later in §IV-A1, these improvements significantly increase the latest spoofing attack capabilities in terms of the spoofed point number, angle coverage, pattern control, and robustness over distance. Following the best practices advocated in [44], we release the raw hardware design file (CAD file) and the bill of materials on our website [45] to make it easier for future works to reproduce and build upon (most recent works did not do this [9, 10, 13]).

IV. OBJECT INJECTION ATTACK MEASUREMENTS

With the measurement setup above, in this paper we perform the first large-scale measurement study for both classes of state-of-the-art LiDAR spoofing attacks: object injection attack and object removal attack, which will be the focus of this and the next section respectively. For each attack class, we first perform the attack capability measurements at the LiDAR point cloud level, and then model such attack capabilities for the subsequent measurements at the object detector level.

A. LiDAR-Level Measurements (RQ1, RQ2)

In this section, we measure the object injection attack capabilities at the LiDAR point cloud level, which is measured by the spoofing attack's capability in injecting spoofed points. We perform this measurement using an attack laser with sufficiently high intensity so that the attack-induced spoofed

points can be easily differentiated from the benign ones. To answer RQ1, we first re-visit the existing state-of-the-art spoofing attacks for point injection on the VLP-16 LiDAR [15], which is predominantly used as the *only* evaluation target in prior works [8]–[10, 12, 13]. After that, we then conduct the measurement study on new-gen LiDARs to answer RQ2.

1) Re-Visiting Design Assumptions Made in Prior Works with VLP-16 (RQ1): Prior works have shown that the number of spoofed points in VLP-16 increases as the quality of the spoofer device improves, ranging from 60 [9] to ∼4k [13]. As shown in Table III, with our spoofer improvements (§III-D), such attack capabilities are further improved significantly, which can now inject >6,131 points indoors and >6,514 points outdoors. This is at least 50% more than those in the latest prior work [13], which capped at ∼4,000. We noticed a concurrent work that may have also made contributions in improving the spoofing capability [46]. The full paper is not available at the time of our writing, but from the abstract it seems that (1) the spoofing capability is still capped at 4200 points, which is still at least 50% fewer than ours; and (2) the study scale is much smaller than ours in terms of LiDAR number (only 2, versus 9 for us) and likely also LiDAR types and generation coverage.

Meanwhile, we also observe that the number of spoofed points and angles in the outdoor setup is generally larger than those in the indoor setup. We consider this reasonable since the legitimate laser reflections are fewer in the outdoor environment (e.g., no wall reflections), leaving more room for spoofing. However, this actually contradicts those reported results in the latest prior work [13]: as shown in Table III, the spoofed points outdoors are significantly fewer than those indoors (1.8k versus 4k) and they attribute this to the lighting condition differences (e.g., they also report that the number of spoofed points decreases from ∼2,500 at night to ∼1,600 at day.) We consider that our results are achieved by our more careful optical setup as described in §III-D since the spoofing laser intensity is much stronger than the sunlight and should not be decayed in a short 10 m flight. We suspect that the optical setup of the previous work is not well calibrated and the laser beam diverges.

> **Finding 1** (RQ1): *With electronic and optical setups, the LiDAR spoofing attack can inject >6,000 points ranging >80°, which is a significantly higher number of points and wider range than in previous studies. Furthermore, previous observations that spoofing distance and lighting conditions can greatly affect the LiDAR spoofing capability [13] no*
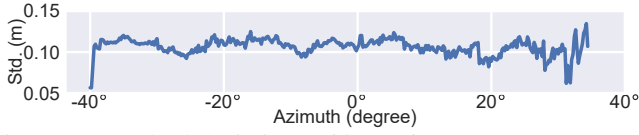
Figure 4: Standard deviations of inner-frame error on VLP-16.

longer hold if implemented using more careful optics.

2) CPI Attack Capability Measurements: As shown in Fig. 1, with our spoofing improvements we are able to demonstrate strong spoofed point pattern control capability for the first time. Considering that the CPI attack capability is a common design-level assumption made in prior works (§II-C2), we thus perform the first systematic quantification of it to allow more rigorous attack capability analysis on the object detector side. Note that some prior works tried to include a modeling of such CPI attack capability in their object detector-side analysis [12], but their modeling is based on their intuitions (e.g., simply add vertical and horizontal noises [12]) instead of the LiDAR sensing mechanisms and the spoofer designs. As shown later in §IV-B3, such erroneous spoofing accuracy modeling can cause significant differences in the object detector-side attack results.

Based on our attack reproduction and spoofing inaccuracy cause investigations, we find that there are two types of errors in controlling the position control of each spoofed point x_{ij} at i-th altitude and j-th azimuth (illustrated in Fig 5): (1) *inner-frame error* $\delta_{ij}^{\mathrm{inner}}$: the inaccuracy in spoofing a point at a chosen 3D position within an individual frame; and (2) *inter-frame error* δ^{inter}: the inaccuracy in maintaining the spoofed position of the same point in the chosen pattern across consecutive frames. Rooted in the spoofing design (Fig. 3), the inner-frame error is due to the inaccurate signal bursting of FG to the gate driver, while the inter-frame one is due to inaccurate timing of triggering the FG from TIA. With our experimental setup, we measure both and find that $\delta_{ij}^{\mathrm{inner}}$ and δ^{inter} are ~10 cm as in Fig 4 and ~35 cm respectively.

Finding 2 (RQ1): *The CPI attack capability, which is commonly assumed but never demonstrated in prior works [9, 10, 12]–[14], can be achievable in practice with a well-calibrated spoofer, while by design such pattern control capability is subject to inner- and inter-frame errors.*

Assumption that object detector-level vulnerability exploitation is needed: Combining these spoofing capability advances above (significantly improved spoofable points, angle, and CPI attack capability), this actually means that *an attacker does not always need to exploit object detector-level vulnerabilities to achieve a near-front road object injection*, which is actually the major design assumption for all attack works so far on the object injection attack side [9, 10, 12]. Specifically, as clearly stated in prior works [9, 10], a valid point cloud for a front-near vehicle needs ~2,000 points and $15°$ azimuth (θ) coverage, but their spoofers can only achieve <200 points and <8°, which is why object detector-level vulnerability exploitation (and thus those novel spoofing pattern optimization/identification) are needed. These results indicate that the exploitation of the object detector-level vulnerability is not the necessary condition although it may help to design more sophisticated attacks.

Table III. Evaluation results of the synchronized injection attack on VLP-16. $\mathcal{N}$: Number of injected points by spoofing. θ: Azimuthal range of the point injection. $\mathcal{R}$: Point injection success rate within θ. The distance d is between the spoofer and the LiDAR. Number in parenthesis: $\mathcal{N}$ from latest prior work [13]. "-": Data not available.

d	Indoor			Outdoor (Daytime: 70 lux)		
	$\mathcal{N}$	$\mathcal{R}$	θ	$\mathcal{N}$	$\mathcal{R}$	θ
2 m	6,523 (-)	98.5%	82.7°	7,705 (-)	94.9%	100.5°
(2.5 m)	(<4k)			(-)		
4 m	6,386 (-)	96.9%	82.5°	7,950 (<1.8k)	96.9%	101.5°
6 m	6,575 (-)	98.6%	83.4°	7,357 (<1.5k)	87.2%	99.6°
8 m	6,213 (-)	93.8%	82.8°	6,702 (<1k)	97.7%	83.4°
10 m	6,131 (-)	93.2%	82.1°	6,514 (<1k)	93.3%	84.2°

Finding 3 (RQ1): *Now with the CPI attack capability demonstrated with sufficiently larger spoofable point number and angle, an attacker actually has sufficient capabilities to directly inject a near-front vehicle pattern. This finding has significant implications for the current line of research, since this (1) may not need to exploit object detector-level vulnerabilities on the attack sides [9, 10, 12]; and (2) may suggest that model-level defenses that try to leverage the inability of spoofers in directly injecting indistinguishable patterns as benign cases (e.g., [10]) can be ineffective.*

3) Measurements on New-Gen LiDARs: Table IV shows the measurement results of the point injection capabilities on not only first-gen LiDARs such as VLP-16 but also new-gen ones. The first thing we find is that the latest spoofing attack for object injection, synchronized spoofing (§II-C1), can only be applied to the first-gen ones; for the new-gen ones, since they all have either timing randomization or pulse fingerprinting, synchronization (and thus CPI) becomes virtually impossible since timing randomization by design prevents the prediction of the future laser firing timing from the victim LiDAR, while pulse fingerprinting by design prevents the prediction of the future laser pulse pattern of each measurement. To still measure their vulnerabilities to spoofing attacks, we try to inject as many spoofed points as possible with a random attack with high-frequency pulses. The random attack is similar to the newly-identified HFR attack (§III-C), but the frequency is tuned to achieve the largest number of injected points.

As shown, compared to the first-gen LiDARs, the security-related features in the new-gen LiDARs result in huge differences in the point injection capabilities: >19k injected points for Horizon, ~3.2k for Helios, 100-350 on Realsense L515 and XT32, and only 28 on OS1-32. We will closely investigate such differences in the next section. Despite these variances, one observation is in common: VLP-16 is actually the *only* LiDAR among all these 7 ones that can achieve the CPI attack capability — it is not possible by design for the 6 new-gen ones due to timing randomization and pulse fingerprinting, while it is also not achievable to the other first-gen LiDAR (VLP-32c) due to simultaneous firing (§IV-A4).

Finding 4 (RQ1): *VLP-16 is actually the only LiDAR for which the CPI attack capability is feasible, which is the key design assumption made in all prior works on object injection attack side [9, 10, 12]. This directly challenges the validity of all these existing designs against the more general and recent set of LiDARs.*

4) Impacts from Security-Related Features:

Table IV. Measurements of the point injection capabilities on different LiDARs. For the first-gen LiDARs, the attack is synchronized spoofing (§II-C1). For the new-gen ones, we inject as many points as possible with random firing (1 MHz). Symbols are the same as in Table III.

	First-Gen		New-Gen				w/ Fingerprint
			w/ Timing Randomization				
	VLP-16	VLP-32c	OS1-32	Helios	Horizon	L515	XT32
$\mathcal{N}$	6,523	9,711	28	3,203	19,182	321	113
$\mathcal{R}$	98.50%	82.90%	43.80%	19.4%	79.90%	0.1%	2.10%
θ	82.7°	73.2°	0.72°	34.2°	103.4°	81.7°	70°

Simultaneous Laser Firing. As listed in Table II, many LiDARs fire multiple laser pulses simultaneously. Velodyne LiDAR is likely adopting a modular approach that doubles units when increasing altitudes. Hence, the number of simultaneous laser firings is also doubled: VLP-16 [15] fires 1 laser during each measurement, and VLS-32 [18] fires 2 lasers simultaneously. This design makes the CPI attack infeasible because we cannot selectively return the laser to each simultaneous laser due to the large diameter of the spoofing laser. For example, on VLP-32c we can always inject spoofed points pair by pair, and the injected pair will always have the same distance to LiDAR. Although simultaneous laser firing may help attackers since it can affect more points by a single attack laser pulse, the CPI becomes no longer feasible.

Timing Randomization. New-gen LiDARs typically have a feature that randomizes their laser shooting timing at each firing to mitigate interference from other LiDARs as listed in Table II. The timing randomization makes the CPI attack capability virtually impossible because the attacker can no longer synchronize with the laser firing pattern. However, interestingly, we find that the attacker may still be able to inject some points if the randomization is not strong enough. To quantify the impacts of the timing randomization, we measure the laser firing interval (not available from data sheets) with a PD and an oscilloscope. Table V illustrates the histogram, standard deviation, and maximum value of laser firing intervals of each LiDAR with timing randomization. We also categorize and fit the observed firing intervals into the uniform or Gaussian distribution based on the shape of its histogram. We convert time difference to distance with the following formula: $\frac{\Delta t \times c}{2}$, where Δt is the timing difference and c is the speed of light. Fig. 16 in Appendix shows the point clouds of Horizon [42], which can illustrate the impacts of timing randomization on spoofing capability. As shown, when under attack, the whole point cloud becomes highly randomized, and the object shapes (e.g., our lab room as shown in the benign case) completely disappeared. We further evaluate the significance of the errors on object detectors in §IV-B3.

Pulse Fingerprinting. We identify that Hesai XT32 [24] can foil synchronization (and CPI attack capability) even without timing randomization due to its pulse fingerprinting. While we cannot be entirely sure due to the lack of official documentation, XT32 likely encodes its fingerprinting in the interval of the pair: As shown in Fig. 6, the XT32 pulse always forms a pair of pulses (two closely-connected spikes) corresponding to a single distance measurement. However, we also find that the fingerprinting itself cannot perfectly defend against spoofing attacks because random spoofing can sometimes coincide with the fingerprinting interval, which can

Table V. Distribution of laser firing intervals for LiDAR with timing randomization. Std. σ: Standard deviation of the error caused by the timing randomization in meters. Max. Δ: Maximum error caused by the timing randomization in meters.

	OS1-32 [22]	Horizon [42]	L515 [41]	Pixell [40]	Helios [23]
Dist. [μs]	$\mathcal{U}_{1.4,1.8}$	$\mathcal{U}_{4.0,4.3}$	$\mathcal{N}_{51,0.025}$	$\mathcal{U}_{4.5,5.8}$	$\mathcal{N}_{1.6,0.005}$
Std. σ	33.3 m	26.0 m	7.5 m	110.4 m	1.5 m
Max. Δ	57.7 m	45.0 m	20.1 m	191.3 m	5.3 m

$\mathcal{U}_{\min,\max}$ - Uniform distribution, $\mathcal{N}_{\mean,\std}$ - Gaussian distribution
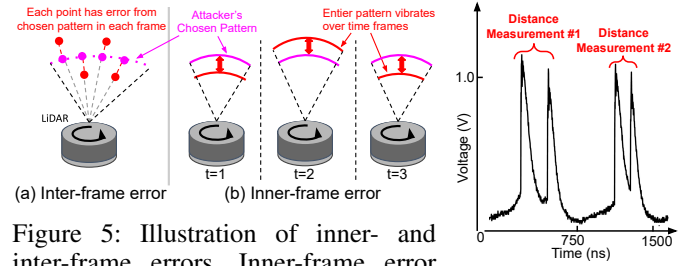

Figure 5: Illustration of inner- and inter-frame errors. Inner-frame error causes spoofing inaccuracy along with the ray direction within a frame. Inter-frame error causes the entire pattern to vibrate across consecutive frames.

Figure 6: Examples of the receiving pulse shape of XT32 [24].

lead to up to 113 spoofed points as in Table IV. This could be because such pulse fingerprinting is originally developed for anti-interference purposes (e.g., to allow multiple LiDARs to operate at close range) instead of security and thus does not have enough randomness/entropy. Furthermore, it is not trivial to design more complex fingerprinting while ensuring eye safety as we will discuss in §VI. This means that if in each attack laser-firing event the attacker also fires a pair of pulses with the interval that can most likely coincide the fingerprinting interval (e.g., the one we found that can spoof 113 points), there can still be a random subset (e.g., 113) of the points in the attacker-chosen point cloud pattern that can be spoofed with the CPI attack capability. Thus, later in §IV-B, we model this effect on the spoofing capability as a random downsampling from a chosen pattern.

> **Finding 5** (RQ2): *The current pulse fingerprinting is not complex enough to perfectly prevent spoofing attacks, likely because it is currently designed only for anti-interference purposes instead of security. However, it is not trivial to design a complex fingerprinting while ensuring eye safety.*

5) Case Study on Specific LiDARs: Using our measurement setup, we also found 5 uniquely-interesting characteristics of specific LiDARs: the use of rare wavelength in OS1-32, relay attack on Leddar Pixell, zero-distance sensing of XT32, extra-wide vertical FOV of Helios 5515, and simultaneous laser firing on OS1-32 and VLS-128, which led to 2 new findings that are different from existing understandings in the community. Details are in Appendix C.

B. Object Detector-Level Measurements (RQ3)

1) Modeling of the Spoofing Attack Capability in Point Injection: Similar to prior works in this problem space [9, 10, 12], to enable large-scale measurements of the attack capabilities on the object detector side, we need to mathematically model the point injection capabilities. Prior efforts in such

modeling did not systematically consider (1) the modeling of the CPI attack capability (§II-C2), and (2) common new-gen LiDAR features that can fundamentally affect the object injection attack capabilities such as timing randomization and pulse fingerprinting. Leveraging this measurement study, we can thus develop a new modeling that addresses both fronts:

$$\mathcal{P}_I(x_{ij}) = x_{ij} + (\delta_{ij}^{\text{rand}} + \delta_{ij}^{\text{inner}} + \delta^{\text{inter}}) \cdot g(x_{ij}), \; x_{ij} \in \mathcal{C}_n \subset \mathcal{C} \quad (1)$$

, where $\mathcal{C}$ is a point cloud (i.e., chosen pattern) that the attacker originally wants to inject (e.g., the point cloud of a vehicle); $\mathcal{C}_n$ is a point cloud randomly downsampled to n points from $\mathcal{C}$ to model the impact from pulse fingerprinting (§IV-A4); $x_{ij} \in \mathbb{R}^3$ is a point injected by the attack at i-th altitude and j-th azimuth; $g : \mathbb{R}^3 \to \mathbb{R}^3$ is a function to obtain a movable unit direction of point x. Due to the physics of LiDAR, each point can only move along with the laser direction. As the LiDAR typically locates at the origin of the point cloud, $g(x_{ij})$ can be written as $\frac{x}{||x||_2}$ in this case. To model the spoofing inaccuracy and the effect of the timing randomization, we add 3 types of errors: $\delta_{ij}^{\text{rand}}$, $\delta_{ij}^{\text{inner}}$, and δ^{inter}. Randomization error $\delta_{ij}^{\text{rand}}$ is the error introduced by the timing randomization. It follows a certain distribution based on the target LiDAR as measured in Table V. We use $\delta_{ij}^{\text{rand}} \sim \mathcal{N}(0, \sigma)$ for the Gaussian distribution and $\delta_{ij}^{\text{rand}} \sim \mathcal{U}(-\frac{\max - \min}{2}, \frac{\max - \min}{2})$ for the uniform distribution. $\delta_{ij}^{\text{inner}}$ and δ^{inter} are the inner-frame and inter-frame errors, respectively. Based on the measurements in §IV-A2, we sample $\delta_{ij}^{\text{inner}}$ from $\mathcal{N}(0, 10 \text{ cm})$ and sample δ^{inter} from $\mathcal{N}(0, 35 \text{ cm})$. Note that $\delta^{\text{inter}} \in \mathbb{R}$ does not depend on each point x_{ij}, i.e., one scalar value is sampled per case and applied to all points in the case. To the best of our knowledge, this is the first modeling of the point injection capability that can cover both first- and new-gen LiDAR features.

Note that this modeling does not cover the impacts from the simultaneous laser firing feature (§IV-A4). This is because although such a feature can make the CPI attack capability infeasible, (1) it cannot prevent the point injection itself; (2) details of the simultaneous firing pattern are not well documented; and (3) the more recent LiDARs (e.g., the ones after 2019 in Table II) do not have such a feature. We thus leave its modeling to future work.

2) Experimental Setups: We perform the experiments based on a scenario in the KITTI dataset. Fig. 17 in Appendix depicts the generated scenario: we place a 3D vehicle object in front of the victim and move the corresponding points to the object's surface, following the same methodology used in [10, 11]. We generate 15 scenarios varying the distance between the victim to the vehicle object from 0 m to 14 m (0 m means the victim's nose touches the vehicle object's tail). Fig. 14 in Appendix shows the number of points on the injected vehicle object. In addition to the original point clouds, we evaluate point clouds downsampled from the original one to 10, 50, 100, and 200 points (i.e., $n = 10, 50, 100,$ and 200 in Eq. 1) as a modeling of the fingerprinting effect (§IV-A4). We judge the object injection as successful if there exists a detected object overlapping with the ground truth area of the injected object, i.e., the IoU between the ground truth and detected object is more than zero.

3) Results:

Impact of Error Modeling on Spoofing Accuracy: We measure the impacts of 3 types of different modeling:
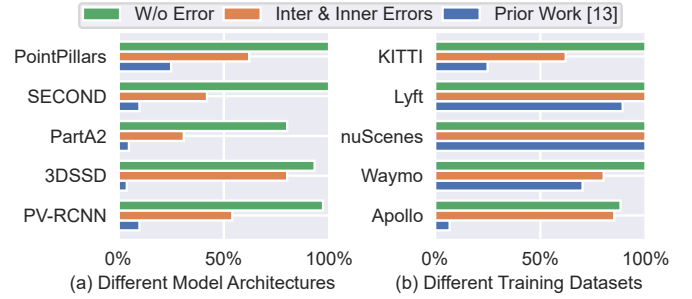

Figure 7: Object injection attack success rates under 3 different types of noises on (a) multiple models trained on the KITTI dataset and (b) PointPillars trained on different datasets.

without errors, our modeling with inner- and inter-frame errors (§IV-A2), and the error modeling used in the Frustum attack [12], which is the latest modeling effort in prior works and simply adds Gaussian errors along with the vehicle's Cartesian coordinates (forward, left, and up) following ($\mathcal{N}(1, 0.1)$, $\mathcal{N}(0, 0.5)$, $\mathcal{N}(1, 0.2)$) in meters, respectively. Fig. 7 shows the attack success rates against our targeted models with different architectures and different training datasets (§III-B). As shown, the different error modeling can generally cause significant differences in the attack success rates, e.g., the results without errors and with the naive modeling in [12] can differ 96% and 70% respectively on average to those using our modeling. Specifically, the latest modeling effort in [12] significantly over-estimates the errors along their forward and up coordinates, which causes the model-level success rates to be generally lower than those with our more systematic modeling based on actual experiments (§IV-A2).

> **Finding 6** (RQ1): *Error modeling on the spoofing accuracy can significantly affect the object detector-level attack results. The latest prior efforts [12] largely overestimate the errors as compared to our systematically modeled and quantified ones via experiments.*

Impact of Pulse Fingerprinting: Fig. 8 shows the attack success rates under different down-sampling levels n as a modelling of different pulse fingerprinting levels (§IV-A4). To perform control experiments of the fingerprinting impact factor, we do not apply inter- and inner-frame errors in this experiment. Generally, fingerprinting with higher complexity (i.e., lower n) shows higher defense capability. However, when $n = 100$, which can represent the fingerprinting complexity level today (XT32 as measured in §IV-A4), the defense effectiveness is actually still very minimal (reduce the attack success rate only by 3% on average). The defense effectiveness only becomes significant when n reaches as low as 10, while such higher effectiveness is still model architecture and training dataset dependent.

> **Finding 7** (RQ3): *If with enough complexity, pulse fingerprinting can indeed show high defense capability to object injection attacks (although can be object detector model architecture and training dataset dependent). Unfortunately, the complexity of pulse fingerprinting today may not be enough to achieve such high defense capability.*

Impact of Timing Randomization: Table VI and VII show the attack success rates under different timing randomization levels based on our measurements in §IV-A4. As shown, timing randomization can in general significantly reduce the
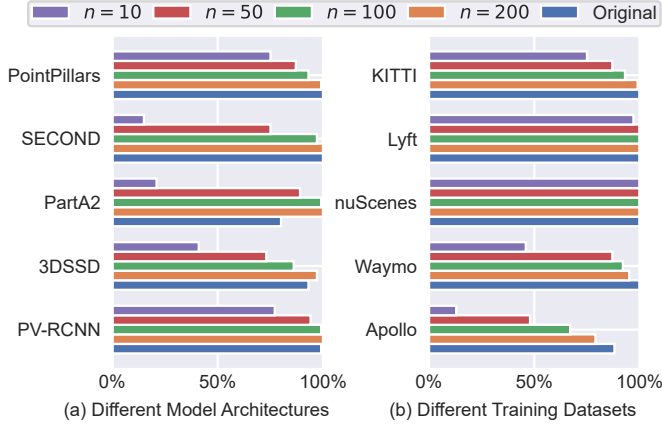
Figure 8: Object injection attack success rates under different down-sampling levels n on (a) multiple models trained on the KITTI dataset and (b) PointPillars trained on different datasets.

attack success rates, e.g., the average attack success rates with randomization are dramatically reduced to at most 34% for most models, while those without randomization are at least 88%. We also measured the attack success rates with both timing randomization and pulse fingerprinting, with the fingerprinting level set at the observed level today ($n = 100$, based on XT32 measurements in §IV-A4). However, consistent with Finding 7, we are not able to observe significant changes in the attack success rate reduction.

> **Finding 8** (RQ3): *Timing randomization, even with those with low randomization entropy that is realized for anti-interference purposes instead of security today, can have significant defense capabilities against object injection attacks. Today's fingerprinting complexity level again may not be able to significantly help boost the defense capabilities when used together with timing randomization.*

Meanwhile, we further notice that the defense effectiveness with timing randomization can differ significantly when applied to models trained on different datasets. For example, as shown in Table VII, the attack success rates are reduced by only 10% and even 0% on average across all randomization levels for the models trained on Lyft and nuScenes respectively, while the success rates can be reduced by as high as 84% and 88.6% for the same model design trained on Waymo dataset and the private dataset from Apollo. In particular, the models trained on Lyft and nuScenes are particularly vulnerable to object injection attacks since as shown in Fig. 8, these models can detect the vehicle even when the vehicle point cloud is randomly downsampled to as few as only 10 points. This suggests that the training dataset choice can play a significant role in determining the model vulnerability to spoofing attacks. Later in §V-B2, we will continue investigating this aspect together with the object removal attack measurement results.

V. OBJECT REMOVAL ATTACK MEASUREMENTS

After the measurements on object injection attacks, in this section we measure the other important class of LiDAR spoofing attacks: object removal attacks (§II-C1).

A. LiDAR-Level Measurements (RQ1, RQ2)

Table VIII shows the LiDAR-level attack capability measurement results for the PRA attack and the newly-identified

Table VI. Object injection attack success rates under different randomization levels based on our measurements in §IV-A4 on *different model architectures*. ∅: No randomization. Avg.: The average results across all these randomization levels.

LiDAR	Rand. model [m]	PointPillars	SECOND	PartA^2	3DSSD	PV-RCNN
VLP-16	∅	100%	100%	80%	93%	97%
Helios	$\mathcal{N}(0, 1.5)$	2%	54%	41%	7%	24%
L515	$\mathcal{N}(0, 7.5)$	0%	24%	14%	7%	0%
Horizon	$\mathcal{U}(-45, 45)$	39%	35%	21%	30%	17%
OS1-32	$\mathcal{U}(-58, 58)$	47%	38%	21%	28%	23%
Pixell	$\mathcal{U}(-191, 191)$	60%	21%	20%	8%	43%
	Avg.	30%	34%	23%	16%	21%
With fingerprinting effect $n = 100$:						
	Avg.	38%	20%	16%	18%	41%

Table VII. Object injection attack success rates under different randomization levels based on our measurements in §IV-A4 on *models trained on different datasets*. ∅: No randomization. Avg.: The average results across all these randomization levels.

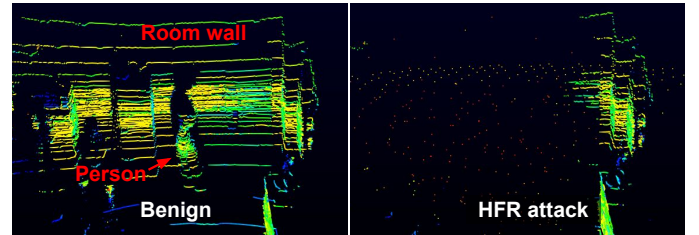| LiDAR	Rand. model [m]	KITTI	Lyft	nuScenes	Waymo	Apollo
VLP-16	∅	100%	100%	100%	100%	88%
Helios	$\mathcal{N}(0, 1.5)$	2%	100%	100%	26%	48%
L515	$\mathcal{N}(0, 7.5)$	0%	60%	100%	0%	0%
Horizon	$\mathcal{U}(-45, 45)$	39%	96%	100%	7%	0%
OS1-32	$\mathcal{U}(-58, 58)$	47%	99%	100%	12%	0%
Pixell	$\mathcal{U}(-191, 191)$	60%	96%	100%	36%	1%
	Avg.	30%	90%	100%	16%	10%
With fingerprinting effect $n = 100$:						
	Avg.	38%	85%	92%	33%	5%

Figure 9: HFR attack effect on VLP-32c. Patterns of a person and the majority of the room wall are completely removed.

HFR attack (§III-C) on different LiDARs. To count the removed points, starting from the attacked point cloud, we first identify the attack-induced spoofed points using the same method as in §IV-A and remove them. Next, we subtract the remaining points in the attacked point cloud from the benign point cloud; now the remaining points in the benign point cloud are thus the removed benign points by the attack (detailed in Appendix F). We select 1 MHz as the pulse frequency and 80V as the laser drive voltage for HFR (detailed in Appendix E).

As shown in Table VIII, with our spoofer improvements in §III-D, PRA can now achieve over 6,600 removed points on VLP-16, which is 65.5% more than the original paper [13]. However, such strong attack capability is limited to the first-gen LiDARs; for the new-gen ones, PRA is not applicable to any of them since its requirement of synchronization is directly foiled by common next-gen LiDAR features such as timing randomization and pulse fingerprinting (§IV-A4).

> **Finding 9** (RQ1): *Due to the requirement of synchronization, the basic design assumption of the latest object removal attack is generally broken for next-gen LiDARs due to common features such as timing randomization and pulse fingerprinting.*

Table VIII. LiDAR-level attack capability measurements for object removal attacks. PRA [13] is only applicable to the first-gen LiDARs. Symbols are the same as in Table III.

| | | First-Gen | | New-Gen | | | | |
| | | | | w/ Timing Randomization | | | | w/ Fingerprint |
		VLP-16	VLP-32c	OS1-32	Helios	Horizon	L515	XT32
PRA [13]	$\mathcal{N}$	6,621	9,711	N/A	N/A	N/A	N/A	N/A
	$\mathcal{R}$	96.9%	82.9%	N/A	N/A	N/A	N/A	N/A
	θ	85.4°	73.2°	N/A	N/A	N/A	N/A	N/A
HFR (§III-C)	$\mathcal{N}$	5,358	8,778	28	4,108	19.2k	206k	113
	$\mathcal{R}$	78.1%	72.2%	43.8%	24.8%	79.9%	91.3%	2.1%
	θ	85.8°	76.0°	0.72°	103.4°	81.7°	70.0°	34.2°

* N/A: Attack is not applicable to the LiDAR

Since there is no need of synchronization, as shown in Table VIII, the newly-identified HFR attack can be generally applied to all LiDARs in the table, no matter if the targeted LiDAR is first- or new-gen. Specifically, for the first-gen ones, HFR is able to achieve comparable point removal capability to PRA, e.g., removing >5,300 points in >85°. Fig. 9 shows an example of the HFR attack effect on VLP-32c. As shown, the point cloud patterns of a person and the majority of the room wall are completely removed, with only some points with random noise patterns left. Compared to PRA, the point removal success rate of HFR is slightly slower (72-78% versus 80-97% in PRA), which is expected since PRA has more precise control of points with synchronization.

For next-gen LiDARs with timing randomization, HFR is still highly-effective since the high-frequency lasers can still hit all legitimate measurements within the affected azimuthal range despite the randomized timing of legitimate laser firing, showing the capability of removing 4k-206k points (for OS1-32 it is only 28 due to the lack of laser diodes with matching wavelength, explained in §IV-A4). For Helios, the point removal success rate is lower mainly due to its extra-wide vertical FOV (Appendix C-D); in the vertical range critical to real-world attack scenarios (e.g., 30-40° in the center), the success rate is similar to others (Fig. 10). For next-gen LiDAR with pulse fingerprinting (XT32), the point removal capability is significantly reduced due to the difficulty for the attack pulses to coincide with the fingerprinting interval. However, as we found in §IV-A4, there are still chances to randomly remove ~100 points by using a pulse frequency that is most likely to coincide with the fingerprinting interval.

> **Finding 10** (RQ2): *For new-gen LiDARs, although they are no longer generally vulnerable to the latest practical attacks such as PRA, they unfortunately still remain generally vulnerable to object removal attacks, with a similar level of practical attack capabilities as PRA, due to the possibility of new asynchronized removal attack designs such as HFR.*

B. Object Detector-Level Measurements (RQ3)

1) Modeling of the Spoofing Attack Capability in Point Removal: Similar to the object injection attack side, we first perform a mathematical modeling of the spoofing attack capabilities in point removal in order to enable large-scale object detection-level attack capability measurements. For removal attacks, the point removal goal is the same for all point measurements that can be hit by the attack laser (e.g., move to within MOT for PRA, or place the point to a random position for HFR). Thus, the main factor that can affect the removal
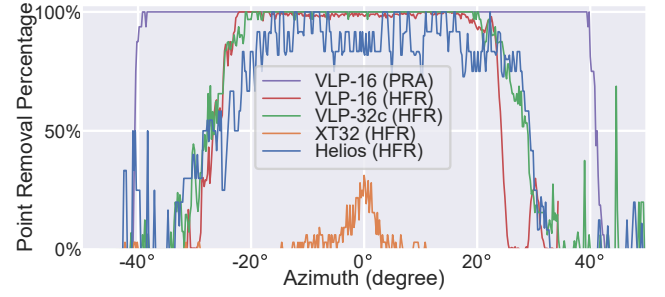

Figure 10: Point removal percentage of PRA [13] and the HFR attack (§III-C) for the azimuth angles under attack.

capability is whether the point is at an azimuth angle that can be effectively hit by the attack laser. For example, Fig. 10 shows the point removal percentage for the azimuth angles under attack. As shown, for VLP-16 under both PRA and HFR, the points in the center of the attack laser-hit azimuth range are almost 100% removed regardless of altitudes and distances, while the removal percentages symmetrically decrease from the center to the side.

Based on these observations, we model the attack capability for point removal $\mathcal{P}_R$ as follows:

$$\mathcal{P}_R(x_{ij}) := \begin{cases} \xi \cdot g(x_{ij}) & \text{if } \text{Bernoulli}(p_j) = 1 \\ x_{ij} & \text{otehwise} \end{cases} \quad (2)$$

, where $x_{ij} \in \mathbb{R}^3$ is its point at i-th altitude and j-th azimuth. For each point, we decide whether it will be removed or not based on a Bernoulli trial with a per-azimuth probability p_j, which can be set using the measured point removal percentage as in Fig. 10. ξ models the removal effects for both HFR and PRA. For HFR, ξ is the error of the HFR attack, which distributes the original points to random locations along with the laser direction $g(x_{ij})$. Thus, ξ can be written as $\xi := \mathcal{U}\left(0, \frac{c}{2} \cdot \frac{1}{f}\right)$, where $\mathcal{U}(a,b)$ is the uniform distribution with the minimum a and maximum b values, c is the speed of light, f is the frequency of HFR attack pulses. This is derived from the fact that the maximum ToF is capped by the time interval of the high-frequency attack pulses in HFR. For PRA, ξ is set to 0 since the point removal principle of PRA is to move the points toward the origin as much as possible to locate them within MOT (§II-C1). After applying Eq. 2, if $\mathcal{P}_R(x_{ij})$ is within the MOT or outside of the maximum detection range (Table II), point x_{ij} is removed from the point cloud. To the best of our knowledge, this is the first mathematical modeling of the point removal capability for LiDAR spoofing attacks.

2) Model-Level Attack Capability Measurement:

Experimental Setup. We use the same evaluation scenarios as in §IV-B2 (i.e., 15 scenarios varying the distance between the victim to the vehicle object from 0m to 14m). We consider the object removal as successful if the IoU between any detected objects and the ground truth object is 0. When applying $\mathcal{P}_R(.)$, we set p_j using the measured point removal percentages for different LiDARs as in Fig. 10. Note that for Helios, since its vertical FOV is much wider than others (Appendix C-D), we only calculate the point removal percentage for the FOV range related to our attack scenario (i.e., 33°, which is enough to cover the height of the front vehicle to remove).

Results. Fig. 11 shows the object removal attack success rates for object detectors with different architectures and datasets. As shown, for the majority of the cases, HFR can reach similar success rates as PRA, which thus further extends our Finding 10 to the object detector level. As a validation of such attack effectiveness in the physical world, Fig. 12 shows the object removal attack effect against real vehicles using the HFR attack. As shown, the HFR attack is found to cause 5 front vehicles to become undetected by the Apollo model [36], with 100% success rate for over 10 seconds. In the figure, we can see the spatial features of the vehicles were completely destroyed (with some random points left) and thus no objects were detectable in the attacked region.

For new-gen LiDAR features, similar to the object injection side, timing randomization (Helios in Fig. 11) can significantly reduce the attack success rate in the majority of the cases (by 35% on average from VLP-16). However, pulse fingerprinting shows much higher defense effectiveness compared to the object injection side: with the pulse fingerprinting strength level of XT32 (i.e., ~100 randomly removed points), the average attack success rate is significantly reduced by 63% on average from VLP-16, while such a reduction is 3% on average on the object injection side (Fig. 8). This is due to the trade-off of the object detector's sensitivity to object injection and removal attacks. Since fingerprinting only allows a random subset of 100 points to be injected, the object detectors that are more vulnerable to object injection are those that are very sensitive to even a heavily-occluded point cloud pattern with a very small number of object points.

Finding 11 (RQ3): *Both timing randomization and pulse fingerprinting show high defense capabilities against object removal attacks in general; the defense capability from pulse fingerprinting is especially strong when compared with that against object injection attacks.*

In addition, similar to our findings on the object injection attack side, the model robustness to object removal attacks shows a high diversity across different model architectures and different training datasets, which is especially prominent across training datasets. Interestingly, the model robustness properties are generally the *opposite* across object injection and removal attacks: for example, the models trained on Waymo and by Apollo are found to be much more robust against object injection attacks when compared to those trained on Lyft and nuScenes (§IV-B), while the former ones become much less robust than the latter ones for object removal attacks. This might be because different dataset has different balances of the trade-off between false positives and more false negatives for the corner cases. As shown in Fig. 8, models trained on Lyft and nuScenes can have 100% detection rate of a vehicle even when the point cloud is randomly downsampled to as few as 10 points. This can allow the trained models to have a very low false negative rate even for highly-occluded legitimate vehicle point cloud patterns, which can thus make the models highly robust to object removal attacks, but this also makes them highly vulnerable to object injection attacks.

Finding 12 (RQ3): *The model robustness to object injection and removal attacks can be highly diverse when trained on different datasets. The choice of training datasets can*
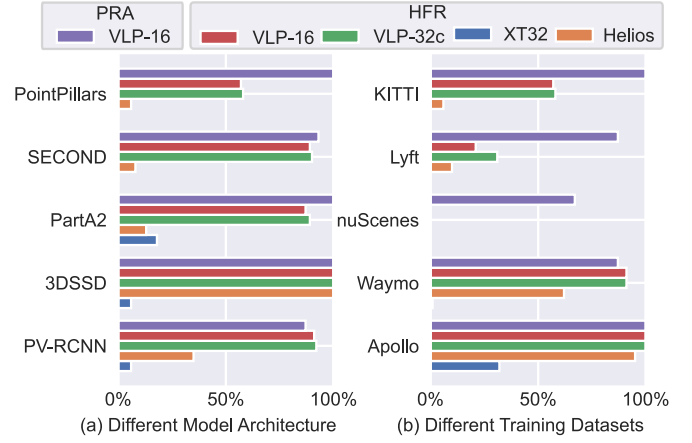

Figure 11: Object removal attack success rates of HFR and PRA for (a) 5 different models trained on the KITTI dataset and (b) PointPillars trained on 5 different datasets.
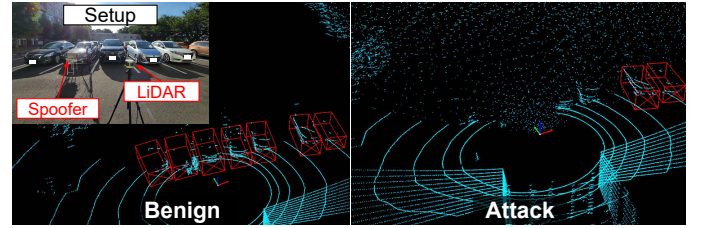

Figure 12: Object removal attack effect against real vehicles using the HFR attack. The 5 front vehicles become undetected with a 100% success rate for over 10 seconds (100 frames in total) by PointPillars [27] in Apollo [36].

---

incur large trade-offs between the model robustness to object injection attacks and that to object removal attacks.

---

3) System-Level Attack Capability Measurement: Due to the downstream components such as object tracking in real-world applications such as autonomous driving (AD), object detector-level object removal effect may not directly imply system-level attack effect such as vehicle crashes [44, 47]. Thus, in this section we further measure the attack capabilities of different object removal attacks at the system level, with the focus on the representative application: AD.

Experimental Setup. We use a common setup widely used in previous work [9, 10, 12, 13]. For the AD system, we use Baidu Apollo 7.0 [36]. For the driving simulator, we use LGSVL [48]. Both systems are known as industry-grade software. Fig. 18 in Appendix illustrates the experiment scenario. We place a sedan vehicle as a target object 200 m away from the victim AD and make the victim AD drive toward the target. We add up to 1 m random perturbation both laterally and longitudinally to (1) the starting point of the victim and (2) the target object. Before reaching the target object, the AD vehicle reaches and keeps 40 km/h. We set an attack start distance $\mathcal{D}$ and only start to apply the simulated removal attack effect when the distance from the victim AD vehicle to the front vehicle is $\leq \mathcal{D}$. We use the collision rate over 10 trials as the evaluation metric, i.e., how many times the victim vehicle collides with the sedan out of 10 trials.

Results. Table IX shows the collision rate over 10 trials for different LiDARs. As shown, although HFR has relatively
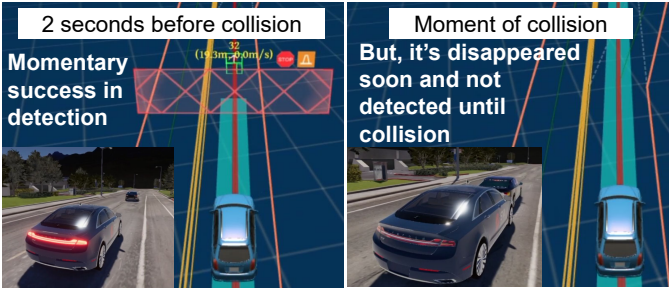
Figure 13: An example HFR attack trial with $\mathcal{D}$ = 15 m on Helios. The remaining points are occasionally detected as an object but are not sufficient to avoid a collision.

Table IX. Vehicle collision rates over 10 trials using PRA and HFR for different LiDARs with varying attack start distances. **Bold** and underline highlight 100% and 0% collision rates.

		Benign	10m	15m	16m	17m	18m	19m	20m
PRA	VLP-16	0/10	0/10	5/10	8/10	9/10	**10/10**	**10/10**	**10/10**
HFR	VLP-16	0/10	0/10	6/10	7/10	8/10	**10/10**	**10/10**	**10/10**
	VLP-32c	0/10	1/10	9/10	8/10	**10/10**	**10/10**	**10/10**	**10/10**
	XT32	0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10
	Helios	0/10	0/10	6/10	5/10	**10/10**	**10/10**	**10/10**	**10/10**

weaker attack capabilities than PRA at raw point cloud (§V-A) and object detector (§V-B2) levels due to the lack of synchronization, their attack capabilities at the system level are almost the same at every attack start distances $\mathcal{D}$, and reach 100% collision rate when $\mathcal{D}$ is over 18 m. However, due to the requirement of synchronization, PRA is only applicable to first-gen LiDARs such as VLP-16, while HRA can be generally applied to new-gen ones such as Helios with similar attack effectiveness. Fig. 13 shows an example attack trial under the HFR attack with $\mathcal{D}$ = 15 m on Helios. The target sedan can be momentarily detected before the collision, but the detection disappears soon and the sedan becomes undetected until and also after the collision.

For next-gen LiDAR with pulse fingerprinting (XT32), although HFR can still have ≤32% model-level attack success rate (§V-B1), it cannot achieve any system-level attack success at all (0%), which is likely because the object removal rate is not high enough to fool object tracking [47]. This suggests that pulse fingerprinting can be quite effective in defending against object removal attacks at the system level. videos of these experiments can be found at our website [45].

> **Finding 13** (RQ3): *Although the HFR attack has weaker attack capabilities than PRA at point cloud and object detector levels, the system-level attack effect is almost the same. Since the HFR attack does not require synchronization, this means that real-world LiDAR applications such as autonomous driving, no matter if using first- or new-gen LiDARs, are facing practical threats from object removal attacks. Meanwhile, pulse fingerprinting can be quite effective in defending against the removal attack effect at the system level.*

VI. Discussions

A. Defense Discussions

1) Sensor-Level Defenses: Through large-scale measurements, we identify the defense capability of popular security-related features in new-gen LiDARs on the spoofing attack capabilities. Table X summarizes our observations regarding their defense effectiveness and limitations against object injection and removal attacks.

Table X. Defense effectiveness and limitations of security-related features in the new-gen LiDARs against object injection and removal attacks. **Bold** means desired properties; Underline means undesire ones.

Features	Effectiveness		Limitations		
	Injection	Removal	Eye safety	Latency	Range↓*
Timing Random.	**High**	**High**	**No risk**	**Low impact**	**None**
Pulse Fingerprint	Mid	**High**	High risk	Mid impact	High
Simul. Firing	Low	None	**Low risk**	**Low impact**	Low

* Range↓: Degradation of the effective sensing range of LiDAR

Defenses against Object Injection Attacks. As shown in §IV-B3, the timing randomization feature has high mitigation capabilities, particularly on several models such as the industry-grade Apollo model. Even with low randomization entropy, the timing randomization shows high defense capabilities. Meanwhile, higher randomization entropy has a higher defense capability. We suggest that the magnitude of the randomization be set as high as possible, e.g., std. dev. $\sigma \geq 0.75$. Pulse fingerprinting also has mitigation capabilities but the complexity of fingerprinting in new-gen LiDARs today is not enough to effectively defend against injection attacks. Finally, simultaneous firing also has a defense effect as it can prevent the CPI attack capability (§IV-A4). However, this alone is unlikely to affect the object injection attack goal in general since the attacker is still able to inject an object point cloud with at least the same number of points as the chosen pattern (but just cannot control the pattern very precisely).

Defenses against Object Removal Attacks. As discussed in §V-B2, both timing randomization and pulse fingerprinting show high mitigation capability against removal attacks, especially pulse fingerprinting in XT32 shows high defense capability even against the HFR attack. The object detectors are typically tuned to avoid false negatives rather than false positives, especially for AD scenarios. This trade-off benefits pulse fingerprinting for defending against object removal attacks because the majority of object points are not compromised. However, this tuning is a backfire to defend against object injection attacks because it allows a hundred injected points to be detected as an object. However, the current XT32-level pulse fingerprinting is still not enough to prevent all attacks as the HFR attack still has 32% attack success rate on the Apollo model (§V-B2).

To improve this further, more complex fingerprint encoding is needed. However, this is fairly non-trivial due to the dilemma between eye safety and detection range. More specifically, the most direct way to increase the fingerprint coding complexity is to increase the number of pulses for each distance measurement. However, the laser power per unit time is capped to ensure eye safety. For example, if we shoot N pulses for each point measurement, the power for each pulse should be $1/N$, and it will roughly degrade the effective sensing range by $1/\sqrt{N}$. To address the dilemma, future work can explore: (1) possible fingerprint coding designs both with high complexity and fewer pulses, and (2) use a wavelength to which the human eye is highly resistant, such as 1550 nm wavelength. For the simultaneous firing feature, for object removal it does not have any defense capability since it just makes the attack effects on some sub-group of the points in the chosen pattern equal instead of having a removal effect of them.

Summary. Overall, the timing randomization and pulse

fingerprinting features show promising defense effectiveness on at least one of the object detector-side attack goals. Particularly, timing randomization has high effectiveness on both attack goals while suffering from the least limitations in eye safety, latency, and sensing range, despite the simplicity of the method. Thus, we strongly recommend implementing it in future LiDARs as a highly cost-effective measure to improve their resiliency against LiDAR spoofing attacks.

2) Software-Level Defenses: Recently, several software-level defenses have been proposed against injection attacks. CARLO [10], SVF [10], and Shadow-Catcher [16] leverage an assumption that the spoofer cannot directly spoof a point cloud pattern with sufficient number of points and precision to make it indistinguishable from a benign one. However, we find that this assumption does not hold since our improved spoofer can achieve the CPI attack capability with a much larger number of points that is sufficient to directly spoof the complete point cloud pattern of a near-front vehicle (§IV-A1). Fortunately, VLP-16 [15] is the only LiDAR for which such CPI attack capability is feasible (§IV-A3). Thus, future research can focus more on more recent LiDARs, especially the new-gen ones, to explore software-level defense design possibilities. For object removal attacks, we still do not have effective defenses since PRA [13] is discovered very recently. A potential software-level defense for the HFR attack is to detect a unique characteristic of the HFR attack that will cause a randomized point distribution pattern (like salt-and-pepper noise) in the area under attack, which is caused by the attack design (§III-C) and can rarely occur in benign cases.

B. Limitations of Our Study

1) Aiming at Driving AD Vehicles: In this study, we do not discuss the deployability of LiDAR spoofing attacks against real-world applications such as AD, especially when the victim AD vehicle is moving in high speed. We focus more on investigating the security properties of different types of LiDARs, rather than the system-level security analysis of autonomous driving in the real world. In a future study, we plan to demonstrate attacks against driving vehicles. A recent study [13] has demonstrated a system capable of victim LiDAR tracking and spoofer aiming to address this problem. Therefore, we expect that targeting a driving vehicle is feasible.

2) LiDAR Model Coverage: While we cover popular LiDARs as many as possible with our best efforts, it is infeasible to cover all public LiDARs due to our budget and their supply capacity. For example, we cannot cover 1550 nm LiDARs such as AEye [49] and Luminar [50] which utilize unique technology, e.g. adaptive scanning and FMCW ToF, and thus potentially have different spoofing capabilities. We also cannot cover private LiDARs such as those in Waymo's robotaxi [2] since they are not publicly available on the market.

C. Considerations for Safe Experiments

All experiments were conducted in a controlled environment and we wore safety goggles for extra eye safety. Note that the unit-area peak power of our laser is actually weaker than prior works [13] as we use a 50% larger aperture lens.

VII. CONCLUSION

In this work, we conduct the first large-scale measurement study on LiDAR spoofing attack capabilities on object detec-

tors with 9 popular LiDARs, covering both first- and new-gen LiDARs, and 3 major types of object detectors trained on 5 different datasets. To facilitate the measurements, we (1) identify spoofer improvements that significantly improve the latest spoofing capability, (2) identify a new synchronized object removal attack (HFR)), and (3) perform novel mathematical modeling for both object injection and removal attacks. Through this study, we are able to uncover a total of 15 novel findings, including not only completely new ones due to the measurement angle novelty, but also many that can directly challenge the latest understandings in this problem space. We discuss defenses based on the findings. We hope that our findings can inspire and facilitate future security research on LiDAR spoofing, especially those targeting safety-critical application domains such as autonomous driving.

REFERENCES

[1] C. Urmson, J. A. Bagnell, C. Baker, M. Hebert, A. Kelly, R. Rajkumar, P. E. Rybski, S. Scherer, R. Simmons, S. Singh *et al.*, "Tartan Racing: A Multi-Modal Approach to the DARPA Urban challenge," 2007.

[2] "Waymo Has Launched its Commercial Self-Driving Service in Phoenix," https://www.businessinsider.com/waymo-one-driverles s-car-service-launches-in-phoenix-arizona-2018-12, 2018.

[3] "Cruise," https://www.getcruise.com/.

[4] "Motional," https://motional.com/.

[5] "Nuro," https://www.nuro.ai/.

[6] "NTSB Investigation Into Deadly Uber Self-Driving Car Crash Reveals Lax Attitude Toward Safety," https://spectrum.ieee.org/ntsb-investigati on-into-deadly-uber-selfdriving-car-crash-reveals-lax-attitude-towar d-safety.

[7] J. Petit, B. Stottelaar, M. Feiri, and F. Kargl, "Remote Attacks on Automated Vehicles Sensors: Experiments on Camera and Lidar," *Black Hat Europe*, vol. 11, p. 2015, 2015.

[8] H. Shin, D. Kim, Y. Kwon, and Y. Kim, "Illusion and Dazzle: Adversarial Optical Channel Exploits Against Lidars for Automotive Applications," in *International Conference on Cryptographic Hardware and Embedded Systems*. Springer, 2017, pp. 445–467.

[9] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial Sensor Attack on Lidar-Based Perception in Autonomous Driving," in *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2019, pp. 2267–2281.

[10] J. Sun, Y. Cao, Q. A. Chen, and Z. M. Mao, "Towards Robust LiDAR-based Perception in Autonomous Driving: General Black-box Adversarial Sensor Attack and Countermeasures," in *USENIX Security Symposium*, 2020.

[11] Y. Cao, N. Wang, C. Xiao, D. Yang, J. Fang, R. Yang, Q. A. Chen, M. Liu, and B. Li, "Invisible for Both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving under Physical-World Attacks," in *IEEE Symposium on Security and Privacy (SP)*, 2021, pp. 176–194.

[12] R. S. Hallyburton, Y. Liu, Y. Cao, Z. M. Mao, and M. Pajic, "Security Analysis of Camera-LiDAR Fusion Against Black-Box Attacks on Autonomous Vehicles," in *USENIX Security Symposium*, 2022.

[13] Y. Cao, S. H. Bhupathiraju, P. Naghavi, T. Sugawara, Z. M. Mao, and S. Rampazzi, "You Can't See Me: Physical Removal Attacks on LiDAR-based Autonomous Vehicles Driving Frameworks," in *USENIX Security Symposium*, 2023.

[14] Z. Hau, K. Co, S. Demetriou, and E. Lupu, "Object Removal Attacks on LiDAR-based 3D Object Detectors," in *NDSS Workshop on Automotive and Autonomous Vehicle Security (AutoSec)*, 2021.

[15] "VLP-16 User Manual," https://velodynelidar.com/wp-content/uploads/2019/12/63-9243-Rev-E-VLP-16-User-Manual.pdf.

[16] Z. Hau, S. Demetriou, L. Muñoz-González, and E. C. Lupu, "Shadow-Catcher: Looking into Shadows to Detect Ghost Objects in Autonomous Vehicle 3D Sensing," in *European Symposium on Research in Computer Security*. Springer, 2021, pp. 691–711.

[17] K. Yoshioka, "A Tutorial and Review of Automobile Direct ToF LiDAR SoCs: Evolution of Next-Generation LiDARs," *IEICE Transactions on Electronics*, vol. E105.C, no. 10, pp. 534–543, 2022.

[18] "Ultra Puck Surround View Lidar Sensor — Velodyne Lidar," https://velodynelidar.com/products/ultra-puck/.

[19] D. Wang, C. Watkins, and H. Xie, "MEMS Mirrors for LiDAR: A Review," *Micromachines*, vol. 11, no. 5, p. 456, 2020.

[20] R. Roriz, J. Cabral, and T. Gomes, "Automotive LiDAR Technology: A Survey," *IEEE TITS*, 2021.

[21] "Ouster Introduces Digital Flash (DF) Series Lidar for Automotive," https://ouster.com/blog/ouster-automotive-df-series/.

[22] "datasheet-rev06-v2p3-os1.pdf," https://data.ouster.io/downloads/datasheets/datasheet-rev06-v2p3-os1.pdf.

[23] "RS-Helios," https://www.robosense.ai/en/rslidar/RS-Helios.

[24] "XT32 - HESAI," https://www.hesaitech.com/en/XT32.

[25] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-Transformed Points," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.

[26] R. Qian, X. Lai, and X. Li, "3D Object Detection for Autonomous Driving: A Survey," *Pattern Recognition*, 2022.

[27] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection from Point Clouds," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 697–12 705.

[28] Y. Yan, Y. Mao, and B. Li, "SECOND: Sparsely Embedded Convolutional Detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[29] S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li, "From Points to Parts: 3D Object Detection from Point Cloud with Part-Aware and Part-Aggregation Network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2647–2664, 2020.

[30] Z. Yang, Y. Sun, S. Liu, and J. Jia, "3DSSD: Point-based 3D Single Stage Object Detector," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 040–11 048.

[31] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10 529–10 538.

[32] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[33] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[34] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[35] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet, "Level 5 Perception Dataset," https://level-5.global/level5/data/, 2019.

[36] "Baidu Apollo," https://github.com/ApolloAuto/apollo.

[37] M. Contributors, "MMDetection3D: OpenMMLab Next-Generation Platform for Feneral 3D Object Detection," https://github.com/open-mmlab/mmdetection3d, 2020.

[38] "OpenPCDet: An Open-source Toolbox for 3D Object Detection from Point Clouds," https://github.com/open-mmlab/OpenPCDet, 2020.

[39] "Alpha Prime," https://velodynelidar.com/products/alpha-prime/.

[40] "Leddar Pixell," https://leddarsensor.com/solutions/leddar-pixell/.

[41] "Intel RealSense LiDAR Camera L515 Datasheet," https://dev.intelrealsense.com/docs/lidar-camera-l515-datasheet.

[42] "Livox Horizon User Manual," https://www.livoxtech.com/3296f540ecf5458a8829e01cf429798e/assets/horizon/Livox%20Horizon%20user%20manual%20v1.0.pdf.

[43] "EOSRAM Radial T1 3/4, SPL PL90_3," https://www.osram.com/ecat/com/en/class_pim_web_catalog_103489/prd_pim_device_2220019/.

[44] J. Shen, N. Wang, Z. Wan, Y. Luo, T. Sato, Z. Hu, X. Zhang, S. Guo, Z. Zhong, K. Li, Z. Zhao, C. Qiao, and Q. A. Chen, "SoK: On the Semantic AI Security in Autonomous Driving," *arXiv preprint arXiv:2203.05314*, 2022.

[45] "Our Project Website," https://sites.google.com/view/cav-sec/new-gen-lidar-sec.

[46] Z. Jin, J. Xiaoyu, Y. Cheng, B. Yang, C. Yan, and W. Xu, "PLA-LiDAR: Physical Laser Attacks against LiDAR-based 3D Object Detection in Autonomous Vehicle," in *IEEE Symposium on Security and Privacy (SP)*, 2023, pp. 710–727.

[47] Y. Jia, Y. Lu, J. Shen, Q. A. Chen, H. Chen, Z. Zhong, and T. Wei, "Fooling Detection Alone is Not Enough: Adversarial Attack Against Multiple Object Tracking," in *International Conference on Learning Representations (ICLR)*, 2020.

[48] "LGSVL Simulator," https://github.com/lgsvl/simulator/.

[49] "AEye Automotive," https://www.aeye.ai/solutions/automotive/.

[50] "Luminar Iris," https://www.luminartech.com/iris/.

[51] "Si PIN photodiode S6775," https://www.hamamatsu.com/us/en/product/optical-sensors/photodiodes/si-photodiodes/S6775.html.

[52] "TL082 LINEAR INTEGRATED CIRCUIT," http://www.unisonic.com.tw/datasheet/TL082.pdf.

[53] "81160A Pulse Function Arbitrary Noise Generator," https://www.keysight.com/us/en/product/81160A/81160a-pulse-function-arbitrary-noise-generator.html.

[54] "EPC9126/EPC9126HC Lidar Demo Boards ," https://epc-co.com/epc/Products/DemoBoards/EPC9126.aspx.

[55] "AFG310 and AFG320 User Manual," https://www.tek.com/en/afg310-manual/afg310-and-afg320-user-manual.

[56] "LiDAR for Automotive and Industrial Applications 2021," https://www.i-micronews.com/products/lidar-for-automotive-and-industrial-applications-2021/.

[57] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud based 3D Object Detection," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4490–4499.

[58] S. Kato, S. Tokunaga, Y. Maruyama, S. Maeda, M. Hirabayashi, Y. Kitsukawa, A. Monrroy, T. Ando, Y. Fujii, and T. Azumi, "Autoware On Board: Enabling Autonomous Vehicles with Embedded Systems," in *ICCPS'18*. IEEE Press, 2018, pp. 287–296.

[59] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 770–779.

[60] R. Girshick, "Fast R-CNN," in *CVPR*, 2015, pp. 1440–1448.

[61] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[62] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot Multibox Detector," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 21–37.

[63] Y. Chen, S. Liu, X. Shen, and J. Jia, "Fast Point R-CNN," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
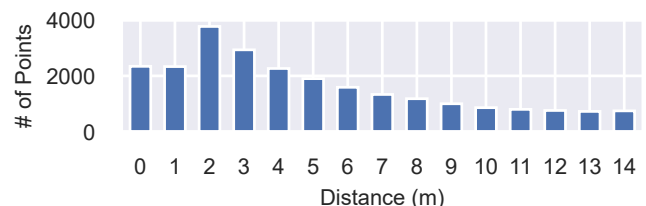
Figure 14: The number of points for the targeted vehicle at each distance to the victim ego vehicle.
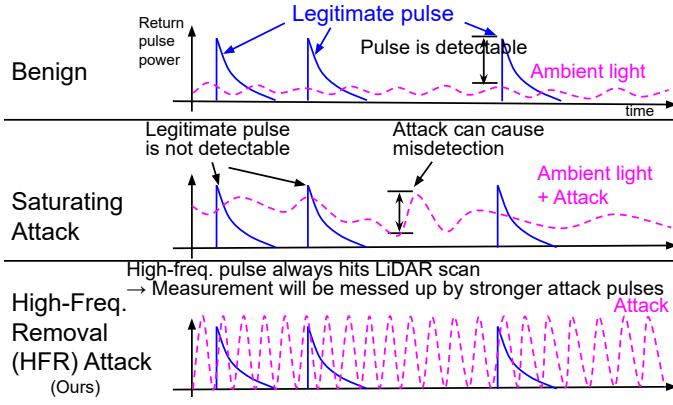
Figure 15: Attack mechanism difference between the saturating attack and the newly-identified HFR attack (§III-C).
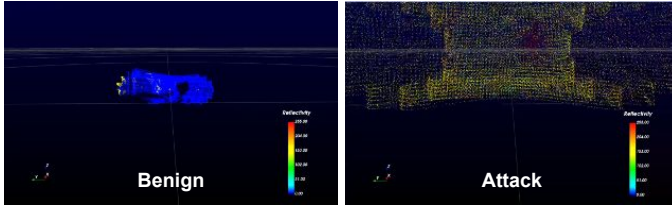

Figure 16: Point cloud of Livox Horizon in benign and attack scenarios. The point cloud is totally randomized by the attack and the pattern of the rectangle area (our lab room) in the left figure completely disappears as shown in the right figure.
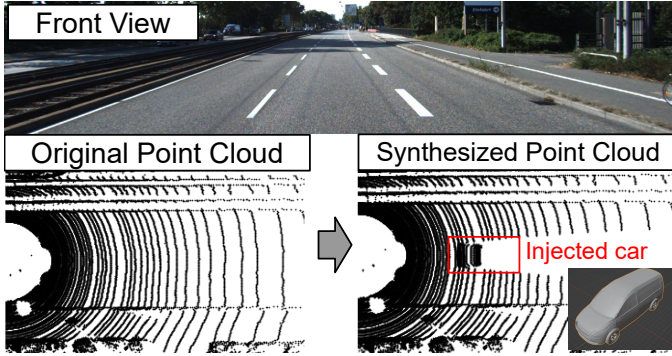

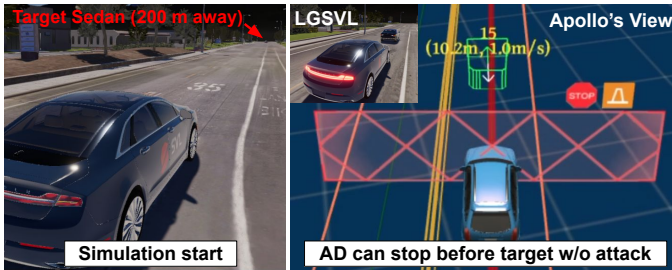Figure 17: Targeted experiment scenario from KITTI [32].

Figure 18: Illustration of the evaluation scenario. AD starts driving from 200 m away and it can always successfully stop before the target sedan in the scenarios without attack.

APPENDIX A
DETAILED EXPLANATIONS ON SYNCHRONIZED LiDAR SPOOFING ATTACKS

Fig. 19 illustrates the synchronized attacks on VLP-16. The attack mechanism is common on both the injection attacks [8]–[10, 12] and the removal attacks [13, 14] since their difference is whether they move points at target locations or move points

into undetectable area. The attack procedure consists of 3 steps: ① PD first receive the legitimate laser from the target LiDAR to know when the LiDAR will scan the point that the attacker want to change; ② FG plans when to fire lasers based on the information from PD and the pre-defined scan pattern of the target LiDAR; ③ the laser is fired based on the plan from FG through the gate driver and LD. Therefore, to achieve the CPI attack capabilities, the attacker must know exactly where LiDAR is scanning and the scan schedule must be predefined or predictable. The timing randomization breaks the assumption by randomizing the scan schedule.
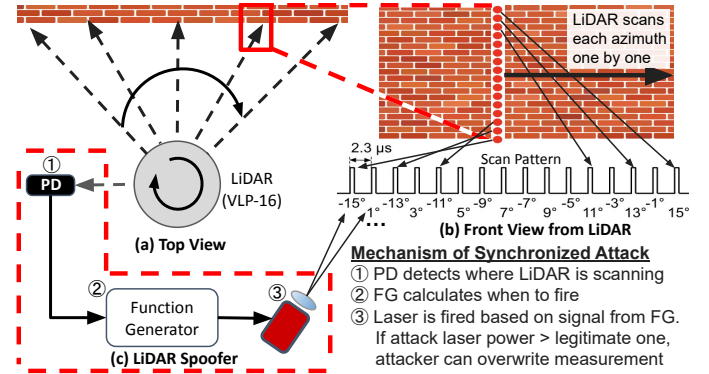

Figure 19: Illustration of synchronized attacks on VLP-16. VLP-16 scans each azimuth (every $0.1°$) one by one. At an azimuth, it fires 16 lasers vertically based on the pre-defined scan pattern. Once attackers can identify its state by PD, they can know *when* to fire a malicious laser based on the pattern.

APPENDIX B
DETAILS OF LiDAR SPOOFER USED IN EXPERIMENTS

Our spoofer operates in 4 steps (shown in Fig. 3): ① The photodetector (PD) receives a laser pulse from LiDAR; ② the transimpedance amplifier (TIA) amplifies the pulse to 100 mV to feed it into the function generator (FG); ③ FG synchronizes with the LiDAR spanning pattern based on the received pulse and generates laser pulses according to the points that the attacker wants to inject; ④ The gate driver (GD) drives the laser diode (LD) which transmits a laser pulse to the LiDAR.

We use the same PD (S6775 [51]) and TIA (TL082 [52]) as the prior work [13]. For other components, we used the equivalent or enhanced devices: FG is Agilent 81160A [53], GD is EPC9126HC [54], and LD is SPL PL90_3 [43]. In particular, we utilize a more functional FG (Agilent 81160A) than the one used in previous studies (Tektronix AFG320 [55]) to better facilitate the exploration of the CPI attack capability.

Furthermore, the improved optical design significantly affects the number and angle coverage of spoofable points as discussed in §III-D. Fig. 20 illustrates the 3 types of laser beam shapes: converged, diverged, and collimated. Among them, the collimated beam is the most ideal for LiDAR spoofing attacks because it can deliver the laser to the target with a minimum loss. To precisely calibrate the lens setup, we develop a device that can adjust the distance between the LD and the lens as designed. As shown in Fig. 3, the lens is connected to the frame with a hollow screw so that we can adjust it precisely.
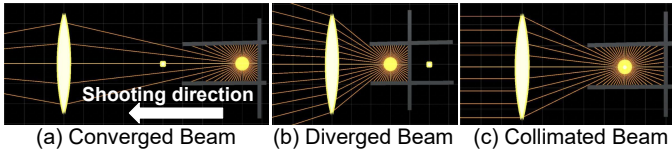
Shooting direction

(a) Converged Beam (b) Diverged Beam (c) Collimated Beam

Figure 20: Illustration of the 3 types of laser beam shapes: converged, diverged, and collimated.

APPENDIX C
CASE STUDY RESULTS ON SPECIFIC LiDARS

A. The Use of Rare Laser Wavelength in OS1-32

For OS1-32 [22], we find that we are not able to inject many points since we only have a 905 nm wavelength (SPL PL90_3), which is also the setup for all prior works [8]–[10, 12, 13]. The spoofing with 905 nm wavelength is not effective on OS1-32 that uses an 865 nm wavelength laser. We tried our best to purchase 865 nm laser diodes, but we find that high-power 865 nm laser diodes are actually not generally publicly available for personal use. We also tried low-power 860 nm laser diodes such as OPV380 and RLD85PZJ4, but they cannot work more than 4 W, which is too low to inject any points (our 905 nm diodes work at 90W to achieve successful injections). From some perspective, this may imply that developing/using a LiDAR with a relatively rare wavelength may have a "immunization" effect against LiDAR spoofing attacks in practice since it may make it harder for attackers to acquire the corresponding attack laser diodes. At this point, only a very small portion of LiDARs has this property; a recent survey [56] reports that 905 nm wavelength laser is the most commonly used and 865 nm wavelength laser does not even appear in their graph due to its rarity.

Nevertheless, such a potential "immunization" effect would lose effectiveness if more LiDARs adopt 865 nm wavelength as this may boost the availability of 865 nm laser diodes on market. Also we find that not all LiDARs using a rare wavelength can directly have such an "immunization" effect. For example, although Realsense L515 is using 860 nm instead of 905 nm, we find that it is still directly attackable with the 905 nm attack laser since it does not have a band-pass filter to exclude light other than its wavelength.

B. Relay Attack on Leddar Pixell

Flash LiDAR fires a wide diverging laser beam and needs to receive them at the same time. This mechanism is exploitable by the relay attack (§II-C1), which is not effective on the scanning LiDARs because PD can only receive a very limited azimuthal range, and the spoofed points will always be located far from the spoofer as discussed in [8]. However, we find that the relay attack can be effective on the flash LiDAR as a removal attack. As its laser covers a wide range, the attacker can also disturb a wider range. If the attacker's laser is strong enough, they can remove the original points by moving them to farther points than the spoofer. Fig. 21 shows the results of the relay attack on Leddar Pixell [40]. The entire row of detection is moved to very far position. Note that Leddar Pixell has a special design that consists of multiple rows of scanning. If it is a typical flash LiDAR, the relay attack can cause an effect on more rows.

Finding 14 (RQ2): *Compared to traditional scanning LiDARs, relay attack may be more effective on flash LiDARs as an object removal attack.*
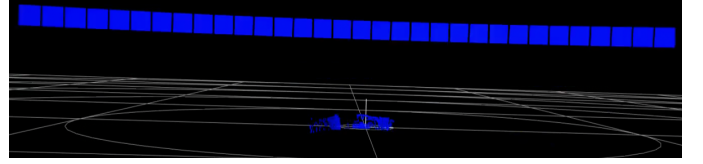

Figure 21: Relay Attack on Leddar Pixell [40]. The attack moves the detected area to a much farther location.

C. Zero-Distance Sensing of XT32

As listed in Table II, XT32 [24] is capable to measure the distance to LiDAR down to 0 m. The zero-distance sensing is not available in most LiDARs due to its technical challenges: detecting a very short time of laser flight is difficult to distinguish from noise due to unideal hardware calibration as discussed in [13]. However, it is technically realizable, which thus directly breaks the design assumption of the latest PRA attack [13] (i.e., needs enough MOT, §II-C1).

Finding 15 (RQ1): *LiDARs with zero-distance sensing do exist, which directly breaks the design assumption of the latest PRA attack [13].*

D. Extra Wide Vertical FOV of Helios 5515

As listed in Table II, Helios [23] has an extra wide vertical field-of-view (FOV), 70°, which is far wider than the normal range of 30-40° (Table II). It results in the lower attack success rate $\mathcal{R} = 19.4\%$ even though the number of points is large, 3,203 points. This is because our spoofer cannot cover all altitudes of Helios. However, it does not mean that Helios is more robust against attacks because the attacker does not need to attack the entire vertical FOV. Fig. 22 shows an example of the HFR attack effect on Helios. As shown, the HFR attack is successful in the middle of the vertical FOV, which can allow attackers to hide objects there.

E. Simultaneous Firing on OS1-32 and VLS-128

Fig. 24 shows the spoofed points on OS1-32 [22] and VLS-128 [39].As shown, OS1-32 fires and scans 32 lasers vertically, and thus we can only move the depth of each vertical line with 32 points simultaneously. VLS-128 shoots 8 lasers based on a predefined pattern, and thus we can only simultaneously change the depth of each group consisting of 8 points. As we do not have the capability to selectively return a laser to each simultaneous laser, we are not able to achieve the CPI attack capability as discussed in §IV-A4.

APPENDIX D
TAXONOMY OF 3D OBJECT DETECTORS

Voxel-based Methods. Voxel-based method is a very early, but still dominant approach [27]–[29, 57]. To deal with the irregular structure of point clouds, this approach aggregates points into 3D voxels to make the CNN work effectively. PointPillars [27] is the most widely used in autonomous driving systems such as Baidu Apollo [36] and Autoware [58] because it can achieve higher throughput by constructing voxels as pillars perpendicular to the grounds although the z-axis
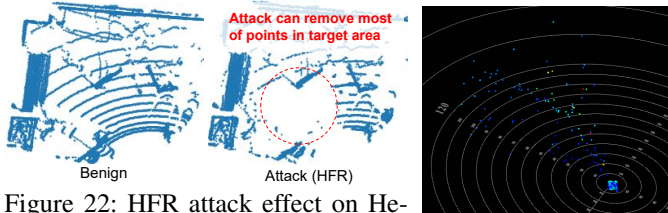
Figure 22: HFR attack effect on Helios [23]. Our spoofer cannot cover the entire FOV, but can still be effective in the center area.
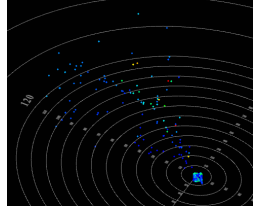

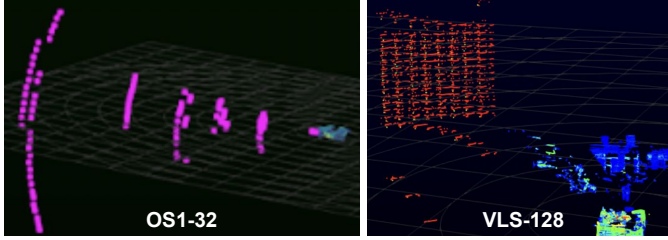Figure 23: Spoofed points on XT32 [24].

Figure 24: Spoofing results for OS1-32 and VLS-128.

Figure 25: The relationship between the attack pulse frequency and the removed points by HFR attack.

resolution becomes coarse. In §IV-B, we find that the voxel-based method is slightly less robust than the other methods to the injection attacks as it cannot recognize the detailed geometry of points. This design allows to achieve of higher throughput and is widely used in autonomous driving systems such as Baidu Apollo [36] and Autoware [58] since the z-axis of an object is typically unimportant in ground vehicles. Part-A^2 [29] achieves a higher attack success rate than the above methods by leveraging multi-scale voxelization and anchor-based strategies.

Point-based Methods. Point-based methods directly handle point cloud without voxelization. To efficiently handle point clouds, this approach utilizes permutation-invariant operators. PointRCNN [59] is a two-stage method inspired by FastRCNN [60] in 2D object detection. PointRCNN generate 3D proposals and point features with PointNet backborns [61]. 3DSSD [30] is a single stage method inspired by SSD [62] in 2D object detection. By leaving only representative points in the downsampling, 3DSSD removes the feature propagation layer and achieves higher throughput than two-stage method.

Point Voxel-based Methods. Point voxel-based method [31, 63] is a hybrid approach of the voxel-based and point-based methods. PV-RCNN [31] is a two-stage method that uses the voxel-based method in the first stage (3D proposal generation) and the point-based method in the second stage (regional refinement). In §IV-B, PV-RCNN [31] shows the highest robustness to the injection attacks. The second stage with PointNet backbone, which is not in 3DSSD, enables to obtain more detailed point geometry.

APPENDIX E
IMPACT OF PULSE FREQUENCY AND LASER DRIVE VOLTAGE ON HFR ATTACK

For HFR attack, the higher the attack pulse frequency is, the more effective the point removal attack capability should be since this can make it more likely for the attack pulse to (1) hit the laser receiving window on the victim LiDAR side to affect the legitimate point measurement; and (2) push the attack-induced random position measurements to be within MOT (and
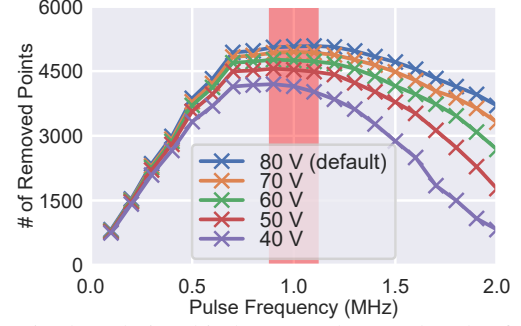
thus become undetectable). However, highly-frequent laser firing makes the temperature of LD high and thus results in the degradation of laser intensity, which affects the spoofing capability. We thus experimentally study this trade-off. As shown in Fig. 25, when we increase the frequency, the number of removed points peaks at ∼1 MHz and decreases after that. Thus, we use 1 MHz as the default attack laser frequency in our other experiments.

Meanwhile, the attack laser intensity at the firing time also practically affects the attack effectiveness, since a lower one is not able to ensure that the attack laser received at the victim is stronger than the legitimate one. To understand this, we also vary the attack laser drive voltage in our experiments. Since VLP-16 has ∼30V laser drive voltage, we vary the attack laser drive voltage from 40 to 80 V (80V is the maximum possible one in our setup). As shown in Fig. 25, the number of removed points monotonically decreases with the voltage value. Thus, we use 80V as the default voltage in our experiments.

APPENDIX F
CRITERIA TO COUNT THE NUMBER OF REMOVED AND INJECTED POINTS

To quantitatively evaluate the attack performance, we design a systematic method to count the number of removed and injected points by attacks. We conduct all indoor experiments 5 m away facing the wall of the room, i.e., the legitimate reflection should be from 5 m away. For point injection attacks, we judge whether a point is spoofed or not based on its intensity. When the intensity scale is 0 to 255, The intensity of reflections from the walls of the room is typically below 70. On the other hand, the attack laser has more than 200 as we directly shoot lasers at the LiDAR without any reflections. We thus used 80 as the threshold for indoor experiments to differentiate whether legitimate or injected points. For the outdoor experiments, we selected an adequate threshold from 80 to 150 based on the maximum intensity in a benign point cloud. For point removal attacks, we also utilize the intensity of the point. We first identify the attack-induced spoofed points with the same threshold of intensity. Next, we subtract the remaining points in the attacked point cloud from the benign point cloud; now the remaining points in the benign point cloud are thus the removed benign points by the attack. To calculate the point removal percentage of each azimuth, we split the area into pies corresponding to each azimuth and apply the same calculation. Based on our measurement, our method can capture 96% of injected or removed points, i.e., 4% of injected or removed points could be missed.

18