WIP: Practical Removal Attacks on LiDAR-based Object Detection in Autonomous Driving

Takami Sato^{*†}, Yuki Hayakawa^{*‡}, Ryo Suzuki^{*‡}, Yohsuke Shiiki^{*‡}, Kentaro Yoshioka[‡], and Qi Alfred Chen[†] [†]University of California, Irvine; [‡]Keio University

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 demonstrates that one can manipulate the LiDAR point cloud and fool object detection by firing malicious lasers against LiDAR. However, these efforts evaluate only a specific LiDAR (VLP-16) and do not consider the state-of-the-art defense mechanisms in the recent LiDARs, so-called next-generation LiDARs. In this WIP work, we report our recent progress in the security analysis of the next-generation LiDARs. We identify a new type of LiDAR spoofing attack applicable to a much more general and recent set of LiDARs. We find that our attack can remove >72% of points in a 10×10 m² area and can remove real vehicles in the physical world. We also discuss our future plans.

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]). While highly beneficial to our everyday life and society, AD is also highly security-critical as even a small operational error can cause fatal consequences [4]. To address this, numerous researchers have been conducting security analyses on Li-DARs [5]-[11] 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 [8], [10], [12] 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 these efforts evaluate only a specific LiDAR (VLP-16) and do not consider the stateof-the-art defense mechanism in the recent LiDARs.

Velodyne VLP-16 [13] 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 [6]. The following works thus all evaluate

Symposium on Vehicles Security and Privacy (VehicleSec) 2023 27 February 2023, San Diego, CA, USA ISBN 1-891562-88-6 https://dx.doi.org/10.14722/vehiclesec.2023.23036 www.ndss-symposium.org TABLE I: Taxonomy of existing LiDAR spoofing attacks and ours. Rows correspond to whether the attack requires the synchronization with the LiDAR scanning pattern. Columns correspond to attack effects: object injection or removal.

	Object Injection Attack	Object Removal Attack
Async. (Black-box)	Relay [5], Saturating [6]	Saturating [6], HFR* (ours)
(White-box)	Adv-LiDAR* [7], Occlusion* [8], Frustum* [10]	PRA* [11], ORA [5]

* Attack effectiveness against AD has been considered.

their attacks only on VLP-16 [7], [8], [10], [11] or use the attack capability on VLP-16 to justify the validity of their threat model [12], [14]. 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 *next-gen*) LiDARs [15], as opposed to the first-generation (or *first-gen*) ones such as VLP-16. The next-gen LiDARs have more advanced spoofing-related features, such as laser timing randomization and pulse fingerprinting. Prior works [6], [7], [11] actually discussed some of them as potential defenses, but none of them actually evaluates the impact and effectiveness of them against LiDAR spoofing attacks.

In this WIP paper, we report our recent progress in designing powerful and practical *asynchronized* (§II-A) spoofing attacks to rigorously measure the vulnerability status of nextgen LiDARs since synchronized ones are directly foiled by their laser timing randomization. Since all existing works only consider first-gen LiDARs, their designs predominately focus on synchronized attacks [6]–[8], [10], [11], leaving the asynchronized attack design space under-explored. To address this, we identify a new asynchronized attack design called *High-Frequency Removal (HFR)* attack, which is much more powerful and practical than prior ones (e.g., can remove points in a 10×10 m² area, while the latest prior one can only remove points in a 41×42 cm² area [6]). We finally discuss the future plans of this research in §V.

II. BACKGROUND AND RELATED WORKS

A. LiDAR Spoofing Attacks

Table I shows a taxonomy of LiDAR spoofing attacks based on (1) the requirement of *synchronization* with the LiDAR scanning pattern (row); and (2) the *attack effect*: object injection or removal (column). The spoofing mechanisms are illustrated in Fig. 1. *Synchronization* means to synchronize the malicious laser firing timing with the victim LiDAR scanning (i.e., laser firing) timing. More details are in Appendix A

Asynchronized Injection and Removal Attack. Relay attack [5] is an asynchronized injection attack, which can inject

^{*}co-first authors



Fig. 1: Illustration of 4 LiDAR spoofing attack types. Synchronized attacks need white-box knowledge of the victim LiDAR scanning patterns and an extra device (PD) for synchronization (§II-A). Asynchronized attacks do not need these (i.e., black-box LiDAR attack), and thus are both more deployable (can work without knowing the victim LiDAR model) and generalizable (to next-gen LiDARs since synchronization is directly foiled by their spoofing-related features). Our HFR attack is the first asynchronized removal attack on par with synchronized removal attacks. PD: Photodetector. Pts: Points.

spoofed points by relaying the laser received from the target LiDAR. It was shown capable of injecting \geq 200 points, but it can only spoof points in farther positions than the spoofer since it needs to first receive a laser pulse before it can send the same pulse back.

Saturating attack [6] is another asynchronized attack that fires a continuous infrared (IR) laser instead of pulsed ones to cause misdetections of laser-receiving events. It is shown that such an attack can inject dozens of randomly placed spoofed points and diminish a 41×42 cm² metal plate. In this work, we are able to identify a new asynchronized attack design that is much more powerful and practical (§III-B).

Synchronized Injection Attack. The synchronized injection attack [6] is proposed to use the *synchronization* described above to overcome the limitations of the relay attack that it can only inject spoofed points farther than the spoofer [5], The early-stage attack designs can inject only 10 spoofed points [6], but the following works progressively increase the number of spoofed points to 60 [7] and 200 [8] and demonstrate that the 60 and 200 spoofed points are enough to cause a false positive, i.e., injecting a fake object. After the success, the attack capability of injecting 200 points becomes the *de facto* threat model in the following works [10], [12], [14].

Synchronized Removal Attack. To remove legitimate objects from detection, 2 synchronized removal attacks have been proposed so far. The first attack is the physical removal attack (PRA) [11] which removes a target object with up to 4,000 points by leveraging the same methodology as the synchronized object injection attack [7]. They utilize LiDAR's minimum operational threshold (MOT), common preliminary filtering to automatically discard points below a certain distance. The second attack is the object removal attack (ORA) [12], which fools 3D object detectors by 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.

B. Spoofing-Related Features in Next-Gen LiDARs.

The advent of first-gen LiDAR has greatly improved AD perception, but its complex mechanical design increases costs and limits scalability. To overcome the limitations, the next-generation (*next-gen*) LiDARs [15] mount all components, such as the photodetector and the readout circuitry, on a single chip known as a system-on-chip (SoC) approach. This approach not only reduces the cost and improves the scalability of the system, but also allows more complex signal processing, such as a large number of simultaneous laser firings [16]–[18], laser timing randomization [18]–[22], and fingerprinting [23]. These features are typically designed to be robust against challenging environments (e.g., multiple LiDARs operating adjacent to each other), but they can also work as a defense or mitigation to spoofing attacks.

Synchronization is no longer possible on Next-Gen LiDAR. Among the new features of next-gen LiDARs, the laser timing randomization makes the synchronized attacks virtually impossible because it can directly foil the assumption of the attack. As discussed in §II-A and Appendix A, the fundamental assumption of the synchronized attacks is the predictability of the scan pattern of LiDAR. However, if the timing of laser firing and receiving is randomized, the attacker can no longer synchronize it or even know when the laser will fire. Due to its simplicity and effectiveness, the majority of next-gen LiDARs [18]–[22] have the timing randomization features. To address the limitation, we design our HFR attack, which does not need the synchronization and thus is robust to the timing randomization by design. We will discuss details of the HFR attack in §III.

C. Threat Model

We follow the same threat model as in prior works [7], [8], [10], i.e., the attacker fires malicious lasers from their spoofer to the victim LiDAR (§II-A, Fig. 1). As described in [10], the spoofer device can be at a front vehicle, vehicle in the next lane, or a roadside in AD scenarios.



Fig. 2: Overview of our LiDAR spoofer setup, the optics design, and the setup of the indoor and outdoor experiments.

III. ATTACK EXPERIMENT SETUP AND METHODOLOGY

Fig. 2 shows an overview of our LiDAR spoofer, its optics design, and the setups of the indoor and outdoor experiments, which will be used in later sections.

A. Our Improvements on Spoofer Design

We generally follow the common setup adopted in the prior works [7], [8], [10], [11], but we improve the electronics and optical setup of the spoofer and achieve a significant improvement enabling the injection and removal of more than 6,000 points, which is 30 times more than the prior injection works [8], [10] and 1.5 times more than the prior removal attack [11]. Particularly, the improvement of the optical setup significantly increases the number of spoofed points even when the distance between the LiDAR and the spoofer is far. To achieve the target LiDAR with a minimum loss, the laser beam should be collimated without diffusion and convergence since even a small misalignment of the lens causes diffusion and convergence and results in the degradation of laser intensity per unit area after a long flight. 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. 2, the lens is connected to the frame with a hollow screw so that we can adjust it precisely.

B. New Asynchronized Removal Attack: High-Frequency Removal (HFR) Attack

As mentioned in §I, to measure the vulnerability status of next-gen LiDARs, we need powerful and practical *asynchronized* attacks since synchronized ones are directly foiled by the timing randomization. In this WIP work, we report our recent progress in designing a new type of asynchronized removal attack called high-frequency removal (HFR) attack, which is illustrated in Fig. 1 and Fig. 6 in Appendix. The key idea of it is to fire a large number of attack laser pulses to the victim LiDAR at a very high *frequency*, which, more specifically, 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 point in that range without any knowledge or synchronization with the victim scanning pattern (i.e., the black-box LiDAR attack model defined in §II-A). However, 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 target objects to a random position or undetectable area of the victim LiDAR (e.g., within MOT). This can completely destruct the point cloud patterns at the original object position, which can thus cause the object removal effects. Moreover, the HFR does not need any feedback from the target LiDAR. The synchronized attacks first need to receive the legitimate laser from LiDAR with PD and thus the attack start timing is limited by the arrival of the laser, which cannot be so strong to ensure human eye safety. On contrary, the HFR attack just needs to aim for the target laser with a high-frequent laser, which can be very strong since the attacker may not care about human eye safety and the laws. The attack effectiveness of the HFR attack mainly depends on how high the attack laser pulse frequency can be; the theoretical attack success rate of the HFR attack can thus be mathematically derived based on the laser frequency.

Comparison with prior removal attacks. Among all spoofing attacks with object removal effect, the latest is PRA, a *synchronized* attack (§II-A). Although it can remove \sim 4,000 points, it requires synchronization and thus compared to HFR, it is by design (1) less deployable due to the white-box attack assumption (§II-A): for HFR, the attack can work without knowing which LiDAR model the victim uses, and can omit the PD part in the spoofer (Fig. 1); and (2) not generalizable to next-gen LiDARs since timing randomization can directly foil synchronization.

On the asynchronized removal attack side, the state-ofthe-art is the saturating attack [6] (§II-A). The fundamental difference is that instead of using *pulsed* lasers to directly manipulate the laser-receiving event timing, the saturating attack works by using a continuous laser to increase the ambient noise level to indirectly cause random detection errors of laserreceiving events, which can thus cause random point injection and removal effects as illustrated in Fig. 6 in Appendix. However, due to the requirement of maintaining continuous high-power laser, it is physically difficult to (1) achieve a large attack laser beam coverage at the victim LiDAR side with sufficiently high intensity, and (2) maintain the attack effect. These thus cause fundamental limitations in the attack capability and practicality, especially when compared to HFR. For example, the demonstrated removal attack effect is only about removing points in a 41×42 cm² area, lasting <4 seconds. On the other hand, our HFR attack leveraging pulsed lasers can remove points in a 10×10 m² area, without any limit on such attack effect duration, which is thus much more powerful and practical, especially for AD settings. Detailed results are shown later in §IV.

IV. EVALUATION

Fig. 3 shows the attack demonstrations of the PRA [11] and our HFR attacks in the indoor setup. We place the spoofer 2 m away from the target LiDAR. As shown, the person and the majority of the room wall are removed by the attacks. For our HFR attack, there are points like a salt-and-pepper noise in the removed area. This is because as the key design feature, the HFR attack is asynchronized and thus achieves removal by moving points to a random location or undetectable area. Table II lists the results of the PRA attack and our newly-identified HFR attack on the first-gen LiDARs and an TABLE II: Evaluation results of the removal attacks (PRA [11] and our HFR attack) on different LiDARs. PRA is only feasible on the first-gen LiDARs (VLP-16 and VLP-32c) due to its reliance on synchronization, which is foiled by the timing randomization. \mathcal{N} is the maximum number of points injected by spoofing. θ is the attacked azimuthal range. \mathcal{R} is the spoofing success rate in the azimuthal range.

-		-		-
		VLP-16	VLP-32c	NG-LiDAR(1)
PRA [11]	\mathcal{N}	6,621	9,711	N/A
w/ our	\mathcal{R}	96.9%	82.9%	N/A
spoofer	θ	85.4°	73.2°	N/A
LIED	\mathcal{N}	5,358	8,778	19,182
	$\mathcal R$	78.1%	72.2%	79.9%
(ours)	θ	85.8°	76.0°	81.7°

N/A: Attack is not applicable on the LiDAR

anonymized next-gen LiDAR (NG-LiDAR(1)) with the timing randomization. Note that we anonymize the next-gen LiDAR for security reasons in this WIP paper.

As shown, due to the reliance on synchronization, PRA is only applicable to the first-gen LiDARs (VLP-16 and VLP-32c); for the next-gen LiDAR (NG-LiDAR(I)), the synchronization is directly foiled by the time randomization as discussed in §II-B. On the other hand, our HFR attack can still be effective on next-gen LiDARs with time randomization, since it does not depend on the synchronization with the fixed scanning pattern. For the first-gen LiDARs, VLP-16 and VLP-32c, we observe the HFR attack has slight attack capability degradation from PRA (e.g., 20% fewer spoofing points for VLP-16), since PRA has more precise control of points by synchronizing with the LiDAR scan pattern. However, we find that such a slight decrease in removal capability does not have significant impacts as shown later in §IV-A.

Impact of Pulse Frequency and Laser Drive Voltage: The attack effectiveness of HFR mainly depends on the attack laser frequency; the higher it is, the more effective the attack should be (§III-B). However, we find that highly-frequent laser firing makes the temperature of LD high and thus results in the degradation of laser intensity, which may affect the spoofing capability. We thus experimentally evaluate this as shown in Fig. 4. As shown, 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 may also practically affect the attack effectiveness, since lower one may not be 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 \sim 30V laser drive voltage, we vary the attack laser drive voltage from 40 to 80 V, where 80V is the maximum possible one in our setup. As shown in Fig. 4, the number of removed points monotonically decreases with the voltage value. Thus, we use 80V as the default voltage in our other experiments.

A. Real Vehicle Removal with HFR attack

We further test the effectiveness of the HFR attack in the physical world. Fig. 5 shows the point clouds in the benign and attack scenarios. We target VLP-16 [13] LiDAR with the dual return mode. We detect objects with the PointPillars [24] model in Baidu Apollo 6.0 [25]. As shown, our HFR attack



Fig. 3: Example results from removal attacks. A person and the majority of the room wall are totally removed by PRA [11] and our HFR attack.



Fig. 4: The relationship between the attack pulse frequency and the removed points by HFR attack.



Fig. 5: Front-vehicle removal attack effect against real vehicles using our HFR attack. The 5 front vehicles become undetected with a 100% success rate over 10 seconds (100 frames in total) by PointPillars [24] in Apollo [25].

is found to successfully remove 5 front vehicles at \sim 5 meters away, out of which all can be correctly detected in the benign scenario. Such an attack effect is found consistent across all the 100 continuous frames we collected, leading to a 100% attack success rate over 10 seconds. In the figure, we can see the spatial features of the objects were completely eliminated (with some random points left) and thus no objects were detected in the attacked region.

V. CONCLUDING REMARKS

In this WIP work, we report our recent progress in the HFR attack, which is the first removal attack that can attack a more general and recent set of LiDARs, which shows high effectiveness in physical-world experiments. In the future, we plan to conduct a large-scale measurement study on LiDAR spoofing attack capabilities on object detectors with multiple next-gen LiDARs. To more rigorously conduct the measurement, we identify the HFR attack to measure the vulnerability status of next-gen LiDARs whose security-related features (e.g., timing randomization) can directly foil the synchronized spoofing attacks. To verify the end-to-end attack effect in AD scenarios, we plan to conduct digital-space experiments with driving simulators and physical-world experiments with driving vehicles. We will also evaluate the defense side and quantize the defense capability of the security-related features in next-gen LiDARs for each spoofing attack including the HFR attack. We will also discuss possible future defenses based on the insights drawn from our measurement.

ACKNOWLEDGEMENTS

This research was supported in part by the NSF CNS-1932464, CNS-1929771, CNS-2145493, USDOT UTC Grant 69A3552047138, JST SPRING JPMJSP2123, JST PRESTO JPMJPR22PA, and JSPS KAKENHI 21K20413.

References

- 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 and it's Called Waymo One," https://www.businessinsider.com/waym o-one-driverless-car-service-launches-in-phoenix-arizona-2018-12.
- [3] "Cruise," https://www.getcruise.com/.
- "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.
- [5] 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.
- [6] 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.
- [7] 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 CCS, 2019.
- [8] 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, 2020.
- [9] 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 S&P*, 2021.
- [10] 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, 2022.
- [11] 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, 2023.
- [12] Z. Hau, K. Co, S. Demetriou, and E. Lupu, "Object Removal Attacks on LiDAR-based 3D Object Detectors," in *AutoSec*, 2021.
- [13] "VLP-16 User Manual," https://velodynelidar.com/wp-content/uploads /2019/12/63-9243-Rev-E-VLP-16-User-Manual.pdf.
- [14] 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.
- [15] 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.
- [16] "Ultra Puck Surround View Lidar Sensor Velodyne Lidar," https: //velodynelidar.com/products/ultra-puck/.
- [17] "Alpha Prime Velodyne Lidar," https://velodynelidar.com/products /alpha-prime/.
- [18] "datasheet-rev06-v2p3-os1.pdf," https://data.ouster.io/downloads/datas heets/datasheet-rev06-v2p3-os1.pdf.
- [19] "RS-Helios," https://www.robosense.ai/en/rslidar/RS-Helios.
- [20] "Livox Horizon User Manual," https://www.livoxtech.com/3296f540ec f5458a8829e01cf429798e/assets/horizon/Livox\%20Horizon\%20user \%20manual\%20v1.0.pdf.

- [21] "Intel RealSense LiDAR Camera L515 Datasheet," https://dev.intelrea lsense.com/docs/lidar-camera-1515-datasheet.
- [22] "Leddar Pixell," https://leddarsensor.com/solutions/leddar-pixell/.
- [23] "XT32 HESAI," https://www.hesaitech.com/en/XT32.
- [24] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection from Point Clouds," in CVPR, 2019.
- [25] "Baidu Apollo," https://github.com/ApolloAuto/apollo.



Fig. 6: Attack mechanism difference between the saturating attack and our HFR attack.

APPENDIX A Detailed Explanations on Synchronized LiDAR Spoofing Attacks

Fig. 7 illustrates the synchronized attacks on VLP-16. The attack mechanism is common on both the injection attacks [6]–[8], [10] and the removal attacks [11], [12] since their difference is whether they move points at target locations or move points into undetectable area. The attack procedure consists of 3 steps as described in Fig. 7. To synchronize with the LiDAR scan pattern, 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.



Fig. 7: 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 scan pattern.