On the Vulnerability of Traffic Light Recognition Systems to Laser Illumination Attacks

Sri Hrushikesh Varma Bhupathiraju
University of Florida
bhupathirajus@ufl.edu

Takami Sato
University of California, Irvine
takamis@uci.edu

Michael Clifford
Toyota InfoTech Labs
michael.clifford@toyota.com

Takeshi Sugawara
The University of Electro-Communications
sugawara@uec.ac.jp

Qi Alfred Chen
University of California, Irvine
alfchen@uci.edu

Sara Rampazzi
University of Florida
srampazzi@ufl.edu

***Abstract*— Connected, autonomous, semi-autonomous, and human-driven vehicles must accurately detect, and adhere, to traffic light signals to ensure safe and efficient traffic flow. Misinterpretation of traffic lights can result in potential safety issues. Recent work demonstrated attacks that projected structured light patterns onto vehicle cameras, causing traffic signal misinterpretation. In this work, we introduce a new physical attack method against traffic light recognition systems that exploits a vulnerability in the physical structure of traffic lights. We observe that when laser light is projected onto traffic lights, it is scattered by reflectors (mirrors) located inside the traffic lights. To a vehicle's camera, the attacker-injected laser light appears to be a genuine light source, resulting in misclassifications by traffic light recognition models. We show that our methodology can induce misclassifications using both visible and invisible light when the traffic light is operational (on) and not operational (off). We present classification results for three state-of-the-art traffic light recognition models and show that this attack can cause misclassification of both red and green traffic light status. Tested on incandescent traffic lights, our attack can be deployed up to 25 meters from the target traffic light. It reaches an attack success rate of 100% in misclassifying green status, and 86% in misclassifying red status, in a controlled, dynamic scenario.**

I. Introduction

Connected autonomous and semi-autonomous vehicles (CAVs) depend on correctly interpreting traffic lights to safely navigate and avoid crashes. Incorrect interpretation of traffic lights can result in critical safety hazards for a victim vehicle, such as sudden brake activation or entry of intersections with active cross traffic. Prior research has focused on traffic light controller security and Traffic Light Recognition (TLR) system manipulation. Attacks on traffic signal controllers typically include network intrusions [1], [2], or traffic congestion-based attacks [3], [4]. These attacks require attackers to compromise traffic light networks or in-vehicle systems.

Recent research explored external attack vectors against TLR systems, such as physical attacks through remote projection of structured light patterns. For instance, the Rolling Colors [5] attack leverages the rolling shutter effect in CAV cameras to alter traffic light classification results. However, this attack requires continuous tracking, aiming, and accurate synchronization with the victim CAV camera's rolling shutter timing. Other research work focuses on improving attack stealthiness. I-Can-See-the-Light (I-C-S-T-L) [6], for example, leverages the susceptibility of certain CAV cameras to infrared (IR) light, which is invisible to humans. In this stealthy attack, errors are induced in the vehicle trajectory. The work also includes building physical structures to create fake traffic lights. However, like Rolling Colors, I-C-S-T-L requires continuously aiming a red LED light at the CAV's camera, and it only targets traffic light detection (rather than classification). Additionally, it is only effective at short distances, even with a high-power light source (up to 10 meters with 30 W IR LED).

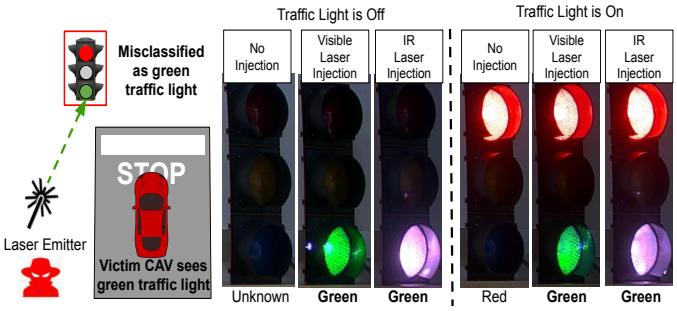


Fig. 1: Overview of our laser injection attack against Traffic Light Recognition Systems (TLRs). The injected laser appears to be a genuine traffic light in CAV camera output, causing misclassification – e.g., a red light misclassified as green.

In this work, we explore a new physical attack against TLR systems that projects light onto the inner physical structure of real-world traffic lights rather than aiming it at the CAV camera sensors. This method avoids the need for synchronization with moving CAVs and does not require the use of advanced skills to hack into vehicle networks and systems, making it more practical than prior work. This attack leverages the unique physical structure of incandescent-based real-world traffic lights. Although new LED-based traffic lights are more power efficient than using incandescent bulbs, updating lights

Symposium on Vehicles Security and Privacy (VehicleSec) 2024
26 February 2024, San Diego, CA, USA
ISBN 979-8-9894372-7-6
https://dx.doi.org/10.14722/vehiclesec.2024.23011
www.ndss-symposium.org

to an array of LED modules remains challenging, requiring several years to complete [7]. Thus, incandescent-based traffic lights are still in widespread use around the world.

Reflectors, such as mirrors, are typically placed within incandescent-based traffic light housings to capture and redirect the light emitted by their internal bulbs. This improves light brightness and visibility, even at a distance [8]. When external laser light is injected, these reflectors scatter the laser light. This light appears to come from the actual traffic light, resulting in potential misclassification by TLR systems (see Fig. 1). More specifically, we demonstrate how this misclassification can be generated by laser light at different wavelengths. For example, a green or red light misclassification can be created by injecting laser light with 532 nm (green) or 650 nm (red) wavelengths, respectively. Similarly, we show that an IR laser light with a 780 nm wavelength, invisible to human eyes, can also be used to cause targeted misclassifications in the TLR model[1]. We design a light injection attack that exploits this structural vulnerability, as instantiated in the reflectors in a real-world incandescent-based traffic light. We evaluate the attack using three industry-grade and state-of-the-art TLR models (Autoware [10], Apollo [11] and LISA [12]). We consider the attack successful when the injection causes the TLR model to misclassify the traffic light (e.g., detect a green status instead of red).

This paper begins with a summary of popular TLR models, traffic light structures, and related work in §II. In §III, we describe our threat model and attack design. We investigate the attack's feasibility with three different laser modules in the visible and invisible spectrum under two test-case scenarios: (i) when the traffic light is *off* (not operational) and (ii) when the traffic light is *on* (operational). We start with feasibility tests in a controlled laboratory setting and find that when the traffic light is off, the traffic light status can be misinterpreted by the victim CAV as red or green status, using both IR and visible laser light injections. When the traffic light is on, our experiments show that both visible and IR light injections can cause misclassification of green and red traffic light status.

In §IV, we characterize the attack parameters and observe that a laser power $\leq$5 mW, similar to that of off-the-shelf laser pointers, is sufficient to cause misclassification of a green or red traffic light status with a visible laser light injection while a power $\leq$100 mW is sufficient for an IR laser light injection. We also find that we can achieve stable misclassification when the laser module is placed up to 25 m from the traffic light. Finally, we evaluate the attack in a proof-of-concept dynamic scenario with the victim camera moving at 2.6 km/h, achieving a 100% attack success rate in misclassifying green status as red and an 86% Attack Success Rate (ASR) in misclassifying red status as green. In summary, our work aims to raise awareness of this new class of vulnerability in traffic lights and TLR models while providing suggestions for potential countermeasures.

II. Background and Related Work

Vision-Based Traffic Light Recognition. TLR models identify and interpret traffic lights, facilitating safe navigation in dynamic traffic environments. For this reason, they are widely adopted in popular CAV systems and frameworks such as Tesla [13], OpenPilot [14], Apollo [11], and Autoware [10]. The traffic light detection pipelines in CAVs typically apply object detection algorithms to camera images to identify potential regions of interest (ROIs) in the image that contain traffic lights [15], [16], [17]. TLR models then use the detected traffic light object to classify the traffic light's status. Our work demonstrates a new attack vectoragainst these TLR systems using light injection.

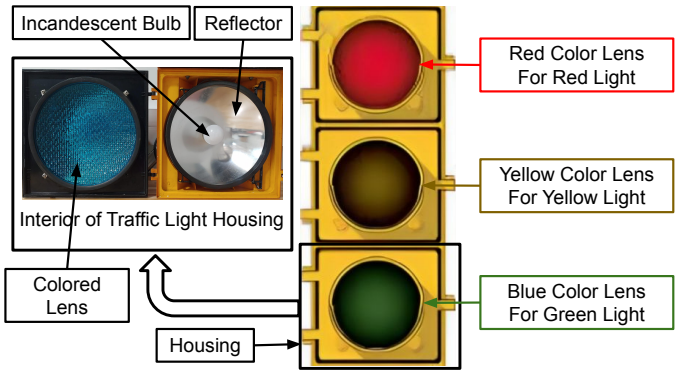


Fig. 2: Structure of the three-light incandescent light-based traffic light. Colored lenses are placed in front of incandescent bulbs to filter and display red, green, and yellow colors.

Traffic Light Colors and Light Sources. Traffic lights come in various types, each serving specific purposes and traffic scenarios. Typical light sources are incandescent bulbs or LED modules that contain an array of LEDs or LED lamps [18]. Although LED-based traffic lights have seen increased usage due to their energy efficiency and longevity, their adoption rate remains low due to the need for infrastructure upgrades, maintenance requirements, associated regulatory approvals, and budget considerations [7], [19]. Thus incandescent-based traffic lights remain widespread in today's transportation infrastructures around the world. An incandescent-based traffic light typically consists of a housing that encloses and protects internal components from weather conditions and other external factors. Incandescent bulbs used inside these traffic lights emit a warm color that is recognized as yellow by human eyes [8]. The required light color (red, yellow, or green) is then achieved by placing colored lenses in front of the bulbs, as shown in Fig. 2. These colored lenses filter the light to display the correct color to drivers. For instance, a blue lens placed in front of an incandescent yellow bulb results in a green color being perceived by human eyes [20]. A reflective material is placed inside the traffic light housing to maximize the visibility of the light emitted by the bulb. When an external laser light source is injected through a colored traffic light lens, the reflector behind the bulb causes the laser light to scatter within the housing. This illuminates the lens, appearing as a genuine light source, as shown in Fig. 1. We leverage this vulnerability in our attack and formulate the threat model discussed in §III. This work considers a standard, incandescent-based, real-world traffic light as shown in Fig. 2.

Prior Attacks using Light Injection. Prior work has shown that light sources can be projected onto a victim camera

[1]Note: IR light appears magenta in images captured by CAV cameras without IR filters, as shown in previous work [6], [9] and Fig. 1

sensor [21], [22], [23], [24], [25], or its surroundings [26], [27], [28], in order to impact machine learning models such as object detectors. For instance, Rolling Colors [5] leverages the rolling shutter properties of some camera sensors to cause traffic light status misclassification. However, this attack requires continuous tracking of the victim vehicle, accurate aiming of the light source, and synchronization with the camera. This can be challenging, given the dynamic nature of vehicles [29]. In this work, we overcome this challenge by exploiting the reflectors inside traffic lights. This makes the attack easier to implement in real-world driving scenarios.

Other work leverages IR light, which is invisible to humans but visible to CAV camera sensors that lack IR filters [6], [9]. For instance, Wang et al. [6] use IR light, invisible to the human eye but perceived by CAV cameras, to create fake traffic lights. The attack requires placing a physical structure that resembles a traffic light near an intersection and then injecting an IR LED light into it to cause the detection of a nonexistent traffic light. These fake structures can be easily recognized by pedestrians or officials while our attack exploits existing traffic lights, presenting a more practical attack scenario. Finally, Sato et al. [9] utilize IR laser light project to perform untargeted attacks on traffic sign classification, leaving the vulnerabilities of traffic light recognition models unexplored.

III. Threat Model and Attack Feasibility

A. Threat Model and Attack Goal

Fig. 1 shows the attack overview. The attacker's goal is to cause targeted misclassification in the victim CAV's TLR model (e.g., confusing green status with red), causing the CAV to behave dangerously, such as unexpectedly stopping or accelerating at an intersection. The attacker achieves this by injecting visible or invisible laser light into the housing of the traffic light. As described in §II, the injected light is scattered upon hitting the reflector in the housing, appearing in the images captured by the CAV camera as a genuine light source.

Attacker's Knowledge and Assumptions. We assume that the attacker can obtain victim CAV camera specifications, such as the presence of an IR filter, by using public information, such as manuals and teardown reports [30], [31]. As considered in the previous work [5], we also assume a black-box oracle access to the TLR model used by the victim CAV. The attacker can use this information to estimate the laser light characteristics required to perform targeted misclassification attack. This information can be found in open source software used by some CAV system developers [11], [10]. We assume that the attacker targets an incandescent traffic light. The attack is remote and requires no access to firmware, or to the images captured by the victim CAV's camera. We also assume that the attacker can choose the laser light wavelength (color). For safety, we assume a maximum optical power of 150 mW for the IR module.

Attack Scenarios. The attacker can place a laser module beside an intersection, with line of sight to the traffic light, to perform stealthy laser injections from up to 25 meters away, as demonstrated in §IV-A. The attacker can perform the attack under two main scenarios: (i) when the traffic light is off (due to a malfunction or power outage) and (ii) when the traffic light is on (functioning properly). The attacker can perform visible light injections, such as the injection of a red light source into the traffic light's red light housing. They can also perform stealthy injections using an IR laser module, which emits light invisible to humans. As discussed in §II, a colored lens is placed in front of the incandescent traffic light bulbs to create a specific color. This colored lens property limits the range of wavelengths the laser module can use for injection. For example, the blue lens only allows green or blue wavelength laser light to pass through. Likewise, the red and yellow lenses allow laser light injection only using yellow and red wavelength range light. In contrast, we observe that IR laser light can be injected into lenses of any color. While IR lasers facilitate stealthy injections, visible laser light gives the attacker the capability to select injection colors based on the intended target misclassification. In light of these findings, our threat model explores both IR and visible light injections.

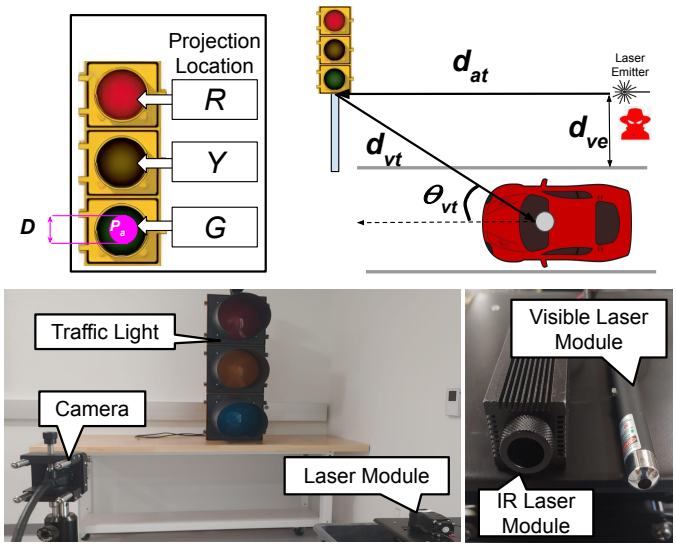


Fig. 3: Overview of variables and parameters of the laser injection attack (top). Controlled experimental setup (bottom-left) and the IR and visible laser modules (bottom-right).

B. Attack Formulation

We formulate our attack as shown in Fig. 3. The attacker considers the following attack parameters to exploit the vulnerability. (1) d_{at}: the distance of the attacker's laser module from the traffic light, (2) d_{ve}: the vertical elevation of the laser module with respect to the ground, (3) P_a: the optical power of the laser beam, (4) D: the laser beam spot diameter at the traffic light, (5) w_a: the wavelength of the laser beam injection, and (6) $\{R, Y, G\}$: the laser spot lens target. For target G, the laser target is the green traffic light lens. Likewise, R indicates a red lens target, and Y indicates a yellow lens target.

We also use the following two parameters related to the victim CAV. (7) d_{vt}: the distance of the victim CAV camera to the traffic light, and (8) θ_{vt}: the angle between the CAV camera and the traffic light. The attacker can select attack parameters that maximize attack effectiveness and stealthiness based on the desired goal, such as misclassifying a red status as green or a green status as red.

TABLE I: Classification results for three TLR models, using IR, and visible laser injection at different spot locations when the traffic light is off. ●: traffic light is off. ●: IR injection [1]. ● and ●: green and red visible laser light injections.

TLR Model	IR Laser Spot Location			Visible Laser Spot Location		
	G	Y	R	G	Y	R
Apollo	Green	Red	Red	Green	Yellow	Red
Autoware	Red-right	Red	Red	Green	Red	Red
LISA	Go	Stop	Stop-left	Go	Stop	Stop

TABLE II: Classification results of the TLR models under IR laser light injection attack when the traffic light is on (operational). ●: traffic light bulb can be on in that location. ●: IR injection[1]. **Bold** texts indicate successful misclassification.

TLR Model	Classification Results w/o injection	Attack Spot Location G	Y	R
Apollo	Green	-	Green	Green
	Yellow	**Black**	-	**Red**
	Red	**Green**	Red	-
Autoware	Green	-	**Red-Right**	**Left-Red-Stg**
	Yellow	Yellow	-	Yellow
	Red	Red	Red	-
LISA	Go	-	**Stop**	**Stop**
	Warning	**Go**	-	**Stop**
	Stop	**Go**	**Go**	-

'-' indicates the attack is not applicable at that location.

C. Methodology and Attack Feasibility

In this section, we investigate the feasibility of attacking TLR models by exploiting the traffic light's physical structure. We consider two types of attack scenarios in our threat model: (i) when the traffic light is off and (ii) when the traffic light is on. When the traffic light is off, the attacker can inject a laser light into any of the three light housings, denoted by $\{R, Y, G\}$, using different laser light wavelengths. When the traffic light is on, the attacker injects the laser light into the remaining two light housings. For example, when the traffic light status is green, the laser light can be injected at the R or Y spot locations. Similarly, when the traffic light status is red or yellow, the laser beam can be injected at the other two spot locations. This creates six different attack variants. We evaluate misclassification under all of these conditions, using three different laser modules: two in the visible range (red and green) and one in the IR range.

Experimental Setup. We perform an attack feasibility test using a real-world incandescent light-based traffic light, as shown in Fig. 2. We use three laser modules, two in the visible range (532 nm green and 650 nm red lasers) [32], and a 780 nm IR laser module from Civil Laser [33]. For the feasibility evaluation, the traffic light is placed at 1 m above the ground. The distance from the laser module to the traffic light d_{at} is set to 1 m, with the laser module placed in front of the traffic light (d_{ve} = 1 m). The victim camera is placed at a fixed distance d_{vt} = 5 m with an orientation $\theta_{vt} = 10°$. This experimental setup is used to replicate a real-world scenario in which the traffic light is on the side of a road, and the victim vehicle is located in the center of a standard 3 m wide lane. The power of the laser modules $P_a = 5$ mW for visible laser light (equivalent power to a laser pointer), and $P_a = 150$ mW for the IR laser module (following the safety constraints). The attack is evaluated in controlled indoor scenarios, with a room ambient light level of 50 Lux. We use an automotive Leopard camera with an OnSemi image sensor as the victim CAV's camera [34] as in previous work [9], [5].

Evaluation Models. We evaluate three popular TLR models: Apollo [11], Autoware [10], and LISA [12]. Apollo uses a CNN-based model for classifying traffic lights as Red, Green, Yellow, and Black classes, where Black means the traffic light is off. Autoware uses a MobileNet [35] architecture to classify traffic lights into 11 different classes, including Red, Green, Yellow, Red-Right, Left-Red-Straight, and Unknown (when the traffic light is off). For the third TLR model, we train a CNN model with an Inception-V3 architecture [36] on the LISA dataset containing US traffic light images. We refer to this model as LISA, which classifies a given traffic light into 7 classes, including Stop, Stop-Left, Warning, and Go.

Results and Observations. Table I shows the classification results of the three TLR models when the traffic light is off. For the IR laser injection at spot location Y, all three models misclassify the traffic light status as red. An IR laser injection at spot position G causes Apollo and LISA models to misclassify the traffic light status as green. When injected at R, Apollo and Autoware models misclassify the traffic light status as red. Finally, when red laser light is injected at spot position R, all three models misclassify the status as red. Likewise, when green laser light is injected at spot position G, all three models misclassify the status as green. Table II and Table III list the classification results for the injection attack with the traffic light on. These results show that misclassification of a red status as green and a green status as red can be achieved in Apollo using visible light injection. The same can be achieved in LISA with IR laser light injection. Although the visible laser can cause misclassification of Go as Stop, it cannot cause misclassification of the Stop class in LISA. In the case of IR laser light injection, Autoware misclassifies the Green class as Red-Right and Left-Red-Straight classes, based on the spot locations in Table II. Autoware does not misclassify yellow and red status for IR laser injections. All three traffic light status (green, yellow, and red) in Autoware can be misclassified as Red-Right with visible laser light injections, as shown in Table III.

IV. Attack Characterization And Evaluation

In this section, we characterize attack capabilities in terms of minimum laser power, attacker location, and the beam diameter required to pursue the attack. We also evaluate a proof-of-concept moving CAV attack scenario.

TABLE III: Classification results of the TLR models under visible laser light injection attack when the traffic light is on (operational). ⚪: traffic light bulb can be on in that location. 🟢 and 🔴: green and red visible laser light injections. **Bold** texts indicate successful misclassification.

TLR Model	Classification Results w/o Injection Attack	Attack Spot Location G	Y	R
Apollo	Green	-	Green	**Red**
	Yellow	Yellow	-	Yellow
	Red	**Green**	Red	-
Autoware	Green	-	**Red-Right**	**Red-Right**
	Yellow	**Red-Right**	-	**Red**
	Red	**Red-Right**	Red	-
LISA	Go	-	**Stop**	**Stop**
	Warning	**Go**	-	Warning
	Stop	Stop	Stop	-

'-' represents that the attack is not applicable at that location.

A. Attack Characterization

For this study, we use the setup discussed in §III. We consider a static scenario where the attacker attempts to cause the most severe result possible, misclassifying green status to red and red status to green, in line with the assumptions in previous work [5]. Examples of successful misclassifications using visible and invisible laser light are shown in Table IV and Table V.

1) Minimum Optical Power: Our feasibility analysis shows that visible light injections require a minimum laser power P_a of 5 mW (the same power as commercial laser pointers used in slide presentations). Thus, here, we measure the minimum power required for the IR module.

Results. Our results for the indoor scenario show that a minimum laser power of 50 mW is required to induce the misclassification of green status in Autoware into Red-Right and Red-Left-Straight classes, as shown in Table II. For LISA, the Go class can be altered to Stop with P_a = 30 mW at spot location R and 60 mW at Y, consistent with our feasibility evaluation results in §III-C. A minimum power (P_a) of 100 mW and 50 mW is required to cause misclassification of a red status as green in Apollo and LISA, respectively.

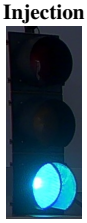
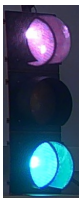
2) Minimum Laser Beam Diameter: We first evaluate the attack with baseline D = 1.5 cm for both IR and visible laser lights. We use the same experimental setup described in §III and a beam expander to increase the beam diameter D from 5 to 25 cm at 5 cm increments. Any further increment ($D > 25$) results in a diameter bigger than the housing of the traffic light.

Results. For IR light, we observe Autoware misclassification of the green status in all beam diameters (1.5–25.0 cm), while LISA requires $D \leq 5$ cm. We achieve red status misclassification for all beam sizes (1.5–25.0 cm) in LISA, $D \leq 5$ cm in Apollo, and only $D = 1.5$ cm in Autoware. The results indicate that Autoware and Apollo require relatively focused IR laser beams (smaller beam diameters, D) to cause

TABLE IV: Classification results of the red status by TLR models under IR and visible laser light injection on a real-world traffic light. **Bold** texts indicate successful misclassification attack results.

TLR Model	No Injection	IR Laser Injection		Green Laser Injection
				
Apollo	Red	Red	**Green**	**Green**
Autoware	Red	Red	Red	**Red-Right**
LISA	Stop	**Go**	**Go**	Stop

TABLE V: Classification results of the green status by TLR models under IR and visible laser light injection on a real-world traffic light. **Bold** texts indicate successful misclassification attack results.

TLR Model	No Injection	IR Laser Injection		Red Laser Injection	
					
Apollo	Green	Green	Green	Green	**Red**
Autoware	Green	**Red-right**	**Left-red-stg**	**Red-Right**	**Red-Right**
LISA	Go	**Stop**	**Stop**	**Stop**	**Stop**

Note that the green light appears blue in cameras without the IR filter.

red status misclassification compared to LISA. Green visible laser light injection causes misclassification of red status as green, regardless of beam diameter, for all three TLR models. Similarly, red visible laser light injection of all beam diameters causes the misclassification of green status as red in all three TLR models.

3) Attacker Position: We characterize the aiming angle, the vertical elevation of the laser module, and the maximum attacker distance to cause misclassification with our setup.

Aiming Angle: Using the experimental setup of §III-C, we aim the laser light at the center of the traffic light's colored lens (aiming angle of the attacker = 0°). We evaluate the injection by varying the aiming angle up to 15° in left, right, upward, and downward directions with respect to the center of the colored lens. We observe misclassification in the three TLR models for all four tested aiming angles. These results indicate that the attack can be successful, regardless of the aiming angle, as long as the laser light is injected into a colored lens, illuminating the internal reflectors.

Vertical Elevation: We set d_{vt} = 2 m, and the traffic

light's elevation to 1 m above the ground. We evaluate the misclassification with the vertical elevation of the attacker laser module, d_{ve} ranging from 0 (ground level) to 1 m (in front of the traffic light) in 4 steps. Our results show that the injection can induce targeted misclassification of all three TLR models in the attack scenarios considered in §IV-A.

Maximum Attacker Distance: We evaluate the maximum attacker distance by setting d_{vt} = 10 m as in §III-C. We set the laser beam diameters at $D = 1.5$ cm and laser power to 150 and 5 mW for IR and visible lasers, respectively. We evaluate the attack on LISA, which shows better performance in classifying green and red status in our controlled scenario. With our experimental setup, laser light injections from up to 25 m from the traffic light succeed in causing red and green status misclassification. This shows that attackers can perform long-range attacks using a simple setup.

Observations. The results indicate that the targeted misclassifications listed in Table §II and Table §III, can be achieved using visible laser light injection with P_a = 5 mW (the power equivalent to a laser pointer for slide shows) and $D \leq 25$ cm or IR laser light injection with P_a = 100 mW and D = 1.5 cm. The results also indicate that the injection can be successfully performed at all the tested aiming angles ($< \pm 15°$), up to 1 m vertical elevations, and with the laser located up to 25 m distance from traffic lights on all three TLR models.

B. Evaluation on Dynamic Scenario

To study the effectiveness of the attack in a realistic setting, we conduct a proof-of-concept evaluation in an indoor dynamic scenario. For this experiment, we place the laser module 5 m from the traffic light. We then record videos, from the victim camera, moving towards the traffic light (along a straight line to simulate a CAV approaching an intersection) from 20 m to 5 m distances at an approximate speed of 2.6 km/h[1]. The measured ambient room lighting is 110 Lux. We test red and green status misclassification in the LISA TLR model using the IR laser module with P_a = 150 mW. The attack is evaluated using ASR, which measures the number of frames in which the attack successfully causes misclassification in the TLR model.

Results. For red status misclassification, we set the spot location of the IR module to G based on our feasibility test. For this case, the attack achieves an 86% ASR for misclassifying the red status to green. We then set the spot location to Y for green status misclassification. Here, the attack achieves a 100% ASR. While this is a proof-of-concept scenario, our results indicate consistent attack effectiveness in dynamic scenarios.

V. Discussion

The attack can cause red and green traffic status misclassification using visible and IR lasers, posing a severe threat to CAV driving frameworks.

We observe that in general, achieving the green status misclassification requires less laser optical power and reduced laser beam focusing (less laser beam convergence) compared to reaching the red status misclassification. We also observe that the tested TLR models might show different robustness to either red or green status classification. For instance, our results in Table II and Table III show that red status classification remains stable in Autoware during the malicious injection, while we achieve green misclassification using both visible and IR laser light injected to both yellow (Y) and red (R) location. On the other hand, Apollo shows robustness for the green status, with the only misclassification happening by direct injection of red visible light in the red (R) location. An attacker can estimate the convenient injection spot and color to perform targeted misclassification by estimating the classification robustness of the target TLR model.

Laser Safety. The experiments in this work were conducted using appropriate eye and skin protection in controlled environments. Note that we set our maximum laser output power to 150 mW (class 3B laser) only for the IR emitter [37].

Potential Mitigations. Our proposed attack injects laser light into the traffic light housing. If the traffic light is already on, two lights might appear simultaneously in the captured images. This is unusual in real-world scenarios and can be used to detect invisible injection attacks. Another potential mitigation technique for IR injection attacks is the detection of anomalous patterns created by IR light. IR laser spots show unique features and texture due to the radiance [38], [9], [39]. A detection technique can be designed to leverage these unusual IR patterns in the captured images to help mitigate the attack.

Limitations and Future Work. Our setup is limited to controlled indoor scenarios and limited speed to meet safety requirements. While we evaluate the attack in indoor lighting conditions (up to 110 lux), linearly higher laser powers might be required to achieve traffic light misclassification outdoors or under high levels of illuminance. Further investigations are needed to evaluate the attack performance, under such conditions. Furthermore, outdoor scenarios are more challenging due to diverse environments and dynamic movement conditions. We leave the outdoor evaluation as future work. Another constraint in performing the injection from far distances is aiming the invisible IR laser at the target spot in the traffic light. However, the attacker can use a proxy visible laser or a camera susceptible to IR light to help aim the laser.

VI. Conclusion

In this work, we explore a new attack vector by demonstrating how internal reflectors of traffic lights can be leveraged to attack industry-grade and state-of-the-art TLR models. We demonstrate that traffic light status misclassification can be achieved using non-sophisticated equipment, such as a laser pointer and an IR laser light with optical power $\leq$ 150 mW more than 25 meters away from the traffic light. Our evaluations in controlled scenarios show the practicality of the attack in low-speed vehicle scenarios reaching up to a 100% attack success rate. Such results show the high risk of compromising CAV safety using existing road infrastructure.

Acknowledgement

We thank the anonymous reviewers for their valuable comments. This research was supported in part by the NSF CNS-1932464, CNS-1929771, CNS-2145493, USDOT UTC Grant 69A3552348327, JST CREST JPMJCR23M4, and unrestricted research funds from Toyota InfoTech Labs.

[1]The limited speed is due to safety and space constraints of our indoor scenario.

REFERENCES

[1] C. Özarpa, İ. Avcı, B. F. Kınacı, S. Arapoğlu, and S. A. Kara, "Cyber attacks on scada based traffic light control systems in the smart cities," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 46, pp. 411–415, 2021.

[2] B. Ghena, W. Beyer, A. Hillaker, J. Pevarnek, and J. A. Halderman, "Green lights forever: Analyzing the security of traffic infrastructure," in *WOOT*, 2014.

[3] Q. A. Chen, Y. Yin, Y. Feng, Z. M. Mao, and H. X. Liu, "Exposing congestion attack on emerging connected vehicle based traffic signal control." in *NDSS*, 2018.

[4] P. Oza, M. Foruhandeh, R. Gerdes, and T. Chantem, "Secure traffic lights: Replay attack detection for model-based smart traffic controllers," in *Proceedings of the second ACM workshop on automotive and aerial vehicle security*, 2020, pp. 5–10.

[5] C. Yan, Z. Xu, Z. Yin, X. Ji, and W. Xu, "Rolling Colors: Adversarial laser exploits against traffic light recognition," in *USENIX Security*, 2022, pp. 1957–1974.

[6] W. Wang, Y. Yao, X. Liu, X. Li, P. Hao, and T. Zhu, "I Can See the Light: Attacks on autonomous vehicles using invisible lights," in *ACM CCS*, 2021, pp. 1930–1944.

[7] J. D. Bullough, J. D. Snyder, A. M. Smith, and T. R. Klein, "Replacement processes for light emitting diode (led) traffic signals," *Contractor's Final Report for NCHRP Project*, pp. 20–07, 2009.

[8] "United States Standard for the Colors of Signal Lights," https://nvlpubs.nist.gov/nistpubs/Legacy/hb/nbshandbook95.pdf, 1964.

[9] T. Sato, S. H. Bhupathiraju, M. Clifford, T. Sugawara, Q. A. Chen, and S. Rampazzi, "Invisible Reflections: Leveraging Infrared Laser Reflections to Target Traffic Sign Perception," in *NDSS*, 2024.

[10] 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*, 2018, pp. 287–296.

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

[12] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Vision for looking at traffic lights: Issues, survey, and perspectives," *IEEE Trans. ITS*, vol. 17, no. 7, pp. 1800–1815, 2016.

[13] Tesla, Inc., "Tesla Autopilot," https://www.tesla.com/autopilot.

[14] comma_ai, "OpenPilot: Open Source Driving Agent," https://github.com/commaai/openpilot.

[15] The Autoware Foundation, "The traffic light map based detector Package," https://autowarefoundation.github.io/autoware.universe/main/perception/traffic_light_map_based_detector/, 2021.

[16] C. Jang, S. Cho, S. Jeong, J. K. Suhr, H. G. Jung, and M. Sunwoo, "Traffic light recognition exploiting map and localization at every stage," *Expert Systems with Applications*, vol. 88, pp. 290–304, 2017.

[17] M. Hirabayashi, A. Sujiwo, A. Monrroy, S. Kato, and M. Edahiro, "Traffic light recognition using high-definition map features," *Robotics and Autonomous Systems*, vol. 111, pp. 62–72, 2019.

[18] Shenzhen Fama Intelligent Equipment CO., Ltd, "the different types of traffic lights that are currently being used," https://www.ledsemaforo.com/news_view-55.html, 2020.

[19] M. Westbrook and W. Rasdorf, "Led traffic signal repair and replacement practices," *Sustainability*, vol. 15, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/1/808

[20] C. Sebes, "Why is there a blue lens on the traffic light?" https://mytrafficlights.com/why-does-the-green-lens-on-my-traffic-light-look-blue/, 2023.

[21] 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, pp. 2267–2281.

[22] 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.

[23] C. Yan, W. Xu, and J. Liu, "Can you trust autonomous cehicles: Contactless attacks against sensors of self-driving vehicle," *DEF CON*, vol. 24, no. 8, p. 109, 2016.

[24] Y. Man, M. Li, and R. Gerdes, "GhostImage: Remote perception attacks against camera-based image classification systems," in *RAID*, 2020.

[25] C. Zhou, Q. Yan, Y. Shi, and L. Sun, "DoubleStar: Long-range attack towards depth estimation based obstacle avoidance in autonomous systems," in *USENIX Security*, 2022, pp. 1885–1902.

[26] G. Lovisotto, H. Turner, I. Sluganovic, M. Strohmeier, and I. Martinovic, "SLAP: Improving physical adversarial examples with short-lived adversarial perturbations," in *USENIX Security*, 2021, pp. 1865–1882.

[27] R. Duan, X. Mao, A. K. Qin, Y. Chen, S. Ye, Y. He, and Y. Yang, "Adversarial Laser Beam: Effective Physical-World Attack to DNNs in a Blink," in *CVPR*, 2021, pp. 16 062–16 071.

[28] B. Nassi, Y. Mirsky, D. Nassi, R. Ben-Netanel, O. Drokin, and Y. Elovici, "Phantom of the ADAS: securing advanced driver-assistance systems from split-second phantom attacks," in *ACM CCS*, 2020, pp. 293–308.

[29] Y. Man, M. Li, and R. Gerdes, "Remote perception attacks against camera-based object recognition systems and countermeasures," *ACM Transactions on Cyber-Physical Systems*, 2023.

[30] J. Yoshida, "Teardown: Lessons Learned From Audi A8," https://www.eetasia.com/teardown-lessons-learned-from-audi-a8/, 2020.

[31] MarkLines Co., Ltd., "BMW 320i Teardown: ADAS/onboard devices," https://www.marklines.com/en/report_all/rep2018_202004, 2020.

[32] "Mini Series Laser Pointer Manual," https://laserpointerstore.com/wp-content/uploads/2021/07/Mini-Series-Laser-Pointer-Manual.pdf, 2021.

[33] CivilLaser, "780nm 1W 2W Powerful IR Laser Module Dot With Cooling Fan," https://www.civillaser.com/index.php?main_page=product_info&products_id=477, 2018.

[34] Leopard Imaging Inc., "LI-USB30-AR023ZWDR data sheet," 2016.

[35] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *IEEE CVPR*, 2016, pp. 2818–2826.

[37] P. J. Smalley, "Laser safety: Risks, hazards, and control measures," *Laser therapy*, vol. 20, no. 2, pp. 95–106, 2011.

[38] J. Goodman, "Statistical properties of laser speckle patterns," *Laser speckle and related phenomena*, pp. 9–75, 1975.

[39] C.-H. Yeh, P.-Y. Sung, C.-H. Kuo, and R.-N. Yeh, "Robust laser speckle recognition system for authenticity identification," *Optics express*, vol. 20, no. 22, pp. 24 382–24 393, 2012.