Fooling Perception via Location: A Case of Region-of-Interest Attacks on Traffic Light Detection in Autonomous Driving

Kanglan Tang, Junjie Shen, Qi Alfred Chen
University of California, Irvine

***Abstract*—The perception module is the key to the security of Autonomous Driving systems. It perceives the environment through sensors to help make safe and correct driving decisions on the road. The localization module is usually considered to be independent of the perception module. However, we discover that the correctness of perception output highly depends on localization due to the widely used Region-of-Interest design adopted in perception. Leveraging this insight, we propose an ROI attack and perform a case study in the traffic light detection in Autonomous Driving systems. We evaluate the ROI attack on a production-grade Autonomous Driving system, named Baidu Apollo, under end-to-end simulation environments. We found our attack is able to make the victim a red light runner or cause denial-of-service with a 100% success rate.**

I. Introduction

In the automotive industry, a revolution is taking place: the rise of Autonomous Vehicles (AVs). AVs have been demonstrated to lower transportation costs and energy consumption, improve travel convenience and comfort, and reduce traffic accidents and congestion [1]. However, the Autonomous Driving (AD) system, which serves as the "brain" of an AV to make driving decisions, may have vulnerabilities that could lead to severe security threats to road safety. For example, vulnerabilities in the localization module, whose outputs are critical for route planning and navigation, can be exploited by attackers to manipulate the routes of AVs [2]. In fact, attacks on the localization will not only affect AVs' route planning and navigation but also have direct effects in other modules such as the perception module to cause false detection.

The perception module processes sensor inputs to perceive the surrounding obstacles and traffic lights, which are necessary for planning a safe driving path and making the correct driving decision. Typically, the perception sensors (i.e., camera, LiDAR, radar, ultrasonic sensor) in the AD system [3], [4] have wide ranges of views. For example, AVs are usually equipped with high-resolution cameras that are capable of detecting obstacles as far as 100 meters or more [3]. However, in this work, we discover an interesting design consideration, which is common in AD systems, that enables an attacker to *blind or misguide the perception without tampering the perception sensor inputs themselves*.

The root cause for such a vulnerability is the design of Region-of-Interests (ROIs). ROI is a strategy commonly employed in AD perception that utilizes information from the localization to *narrow* the detection scope in the sensor inputs [3], [4]. It can effectively reduce the computation overhead by running detection algorithms on smaller input dimensions and filter detection noises by preventing ambiguous detection, e.g., when multiple traffic lights exist in the camera view. Despite the benefits in improving the detection efficiency and accuracy, the accuracy of ROI mainly depends on the localization results–a wrong localization would result into a wrong ROI, which in turn causes the perception module to look at a wrong area in the sensor input.

For AD systems, cameras are especially critical for traffic light (TL) detection since they are the only sensors that are able to accurately detect TL colors. Thus, as the first study, we start from the TL detection and leverage the existing GPS spoofing attacks [5], [6] on localization to demonstrate attack consequences of such ROI attacks in an end-to-end AD system. Attack demos showing the end-to-end attack consequences are at **https://sites.google.com/view/roiattack**.

Considering the severity of the attack, we hope this work can bring immediate attention to the AD system developers for more robust ROI designs. In summary, while this work is still work-in-progress, it makes the following contributions:

- We perform the first security analysis on the ROI design in the perception module in AD systems and identify a design-level vulnerability that allows the attacker to fool AD perception using GPS spoofing.
- We design a concrete ROI attack targeting the TL detection in AD perception. Results show that our attack is able to achieve a 100% success rate in causing the victim AV to run red lights or denial-of-service.

II. Background

A. AD Systems and TL Detection

Baidu Apollo overview. Baidu Apollo [3] is a production-grade open-source AD system for Level-4 AVs [7], which has already been deployed for RoboTaxi services in China [8]. It follows a typical Level-4 AD system design, which mainly consists of localization, perception, prediction, planning, and control modules [3]: the localization module localizes the AV on a map; the perception module uses perception sensors to detect obstacles and TLs; the prediction module predicts the future trajectories of the obstacles; the planning module incorporates maps, localization outputs, and predicted obstacle trajectories to calculate a safe driving trajectory; the control module executes the planned trajectory by actuating the steering, throttle, and brake on the AV. In this work, we focus on two of these modules: localization and perception.

The localization module takes inputs from sensors such as GPS and LiDAR to estimate the real-time position of the AV.

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2021
21 February 2021
ISBN 1-891562-68-1
https://dx.doi.org/10.14722/autosec.2021.23029
www.ndss-symposium.org

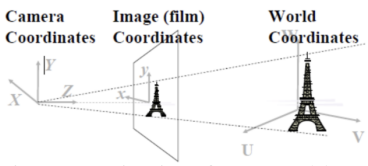

Fig. 1: Projection from world coordinates to image coordinates.

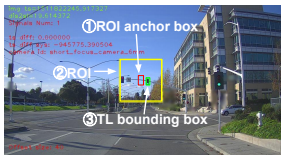

Fig. 2: Illustration of the ROI in TL detection.

Depending on the sensor availability, the localization module can be configured in GPS mode or Multi-Sensor Fusion (MSF) mode [9]. The GPS mode directly takes positioning from GPS, while the MSF mode combines GPS and LiDAR for more accurate and robust positioning. For example, Baidu Apollo, by default, uses MSF in the localization; however, it also provides a GPS mode when there is no LiDAR installed on the AV. Tesla Autopilot, as a Level-2 AD system [7], also do not use LiDARs and thus relies on GPS for localization [10].

The perception module incorporates sensors such as cameras, radars, LiDARs, and ultrasonic sensors to recognize vehicles and pedestrians surrounding the AV. In addition, it also performs important tasks such as TL detection.

TL detection and ROI design. TL detection is an essential feature for Level-4 AVs. If an AV cannot detect and recognize TLs, it may violate traffic rules resulting in some catastrophic consequences, such as running red lights and causing car accidents. Recently, as Level-2 AV companies aiming to achieve higher level driving autonomy, companies such as Tesla starts to add TL detection into their AD systems [11]. TL detection leverages Deep Neural Networks (DNNs) to detect and classify TLs in the camera images [3], [4]. For example, Baidu Apollo applies two DNNs for TL detection; one for the object detection to recognize the TL object in the image, and the other for the classification to recognize the light color [3].

Since the camera image may contain multiple TLs, it is thus necessary to identify the correct TL for the current intersection. To address this, AD systems commonly adopt an ROI design [3], [4], which projects the current TL in the world coordinates obtained from the High-Definition (HD) map to image coordinates based on the localization outputs (Fig. 1). To compensate for any localization or HD map inaccuracies, an ROI area with an empirically determined radius or height/width centered at the projected image coordinates is selected for TL detection. Such an ROI design not only helps prevent ambiguous TL detection but also reduces the computation overhead. Fig. 2 show an example of the ROI in Baidu Apollo. As shown, the ROI anchor box is the projected position of the TL. However, due to inaccuracies in localization and HD map, the projected position is not perfectly aligned with the actual TL. Thus, Baidu Apollo defines a larger rectangle as the ROI for TL detection (② in Fig. 2). After that, it crops the ROI area from the image and applies DNNs to recognize all TL bounding boxes and their colors in the ROI. Finally, the TL bounding box (③ in Fig. 2), which is the closest to the ROI anchor box, is selected as the TL detection result.

B. GPS spoofing attacks to AD localization

Civilian GPS systems are known to be vulnerable to spoofing attacks due to the lack of signal authentication in the infrastructure. In the spoofing attack, the attacker broadcasts fake satellite signals to the victim GPS receiver to deceive it into resolving false positions specified by the attacker [5]. So far, GPS spoofing has been demonstrated to be feasible on various systems, including smartphones [12], drones [5], and vehicles such as Tesla [10].

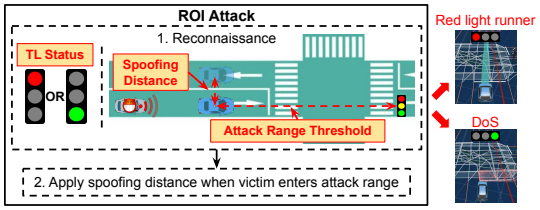

Fig. 3: Illustration of the ROI attack for TL detection.

The attack capabilities of GPS spoofing are different in GPS-based localization and MSF-based localization. In GPS mode, GPS inputs are the only positioning source for localization, and thus GPS spoofing can directly control the localization outputs in the AD system. In MSF mode, since the positioning from GPS will be fused with LiDAR, GPS spoofing cannot set an arbitrary position in the localization output. Nevertheless, Shen et al. [6] have demonstrated that they can inject a large deviation in the MSF-based localization outputs using GPS spoofing.

III. Threat Model and Attack Goal

Threat model. Similar to the threat model used in prior works [6], [12], we assume a car-following model where the attacker launches GPS spoofing attack while tailgating the victim AV. We assume the attacker can arbitrarily manipulate the victim's localization outputs under GPS spoofing, which is valid for AVs using only GPS for localization, such as Baidu Apollo in GPS mode and Tesla. We also assume the attacker can track the physical position of the victim AV, which is feasible if the attacker's vehicle is also an AV [6].

Attack goal. The goal of our attack is to influence the position of ROI in the perception module via GPS spoofing, and thus cause the victim AV to detect a wrong TL to run the red light or fail to detect any TLs to stop at the green light (i.e., denial-of-service or DoS).

IV. Attack Insight and Design

Attack insight. Although from high-level design, the localization and perception modules appear to be independent of each other, the perception module in fact heavily relies on the localization, especially for the ROI logic in TL detection as mentioned in §II-A. Since the localization output directly determines the ROI that the TL detection pipeline will be performed on, any errors in the localization would result in a wrong ROI and may cause incorrect TL detection.

Attack design. To perform the ROI attack on TL detection, the attacker launches GPS spoofing when the victim AV is in front of an intersection. When the victim AV drives towards the intersection, a longitudinal error in the localization would naturally have a smaller effect on the ROI position compared to a lateral deviation. Thus, to maximize the effect on the ROI, we design the attack as spoofing a *lateral* distance away from the physical position of the victim AV, e.g., the spoofing distance shown on the Fig. 3.

Fig. 3 illustrates the ROI attack for TL detection, which is composed of a reconnaissance stage and a spoofing stage. In the reconnaissance stage, the attacker examines the current TL

status and determines the best attack parameters to be applied. Depending on the TL status, the attacker will use different spoofing distances to achieve the desired attack goals. For example, when the TL is red, the attacker needs to spoof a distance such that the ROI includes another green TL in the camera view to cause false detection. Alternatively, if the TL is green, the attacker can simply spoof a distance large enough such that the ROI contains no TLs, which would trigger a safe stop in the planning since the AV knows from the HD map that there is a TL in the front but fails to detect one. In addition, we set an *attack range threshold* for determining the timing to launch the spoofing. If the attacker starts spoofing when the victim is too far or too close to the intersection, the victim might have already been deviated out of the road boundary that prevented it from reaching the intersection or have already detected the correct TL color and committed the correct driving decision. After that, the attacker closely monitors the victim's position and spoofs the corresponding lateral distance when the victim enters the attack range.

V. Evaluation

Experimental setup. Due to the high cost of evaluating self-driving algorithms on real AVs, we follow the common practice for AV testing and perform a simulation-based evaluation, in which we run Baidu Apollo in an industry-grade AV simulator, LGSVL [13]. To facilitate AV simulation, LGSVL provides photo-realistic simulation environments with diverse road structures and a wide range of vehicle models. In our evaluation, we use the Shalun map and Lincoln MKZ vehicle. The Shalun map models a common two-lane road with multiple intersections along the road. We use Lincoln MKZ since it contains the compatible sensor configurations for Baidu Apollo. We simulate our attack on Baidu Apollo version 5.0, which is the latest version fully supported by LGSVL.

To simulate the attack consequences, we create two concrete attack scenarios: *red light detection* and *green light detection*. Specifically, for red light detection, we set the front TL in the first intersection to red and the back TL in the second intersection to green; for green light detection, we set the front TL to green and the back TL to red. We have confirmed that in benign driving, the AV will always correctly detect the front TLs, and make the correct driving decisions, i.e., stop before the intersection or drive through the intersection.

Evaluation metric. In our evaluation, we explore the attack effectiveness of the ROI attack under different attack range thresholds and spoofing distances, aiming to find the parameters that can achieve the highest attack effectiveness. We define a successful attack case as the victim AV mis-detects the front TL and commits the wrong driving decisions, i.e., running the red light or stopping in front of the green light. Since both Baidu Apollo and LGSVL involve random factors such as messaging delays, we calculate a success rate by repeating the simulation for 5 times for each attack parameter.

Attack construction. For the ease of evaluation, we implement the attack logic as an independent module in Baidu Apollo to receive original GPS inputs from LGSVL. If the victim AV is within the attack range threshold, we apply the spoofing distance to the original GPS positions. After that, we publish the spoofed GPS inputs to the localization module.

Results. As mentioned in §IV, the red light detection scenario has a more restricted spoofing distance requirement since

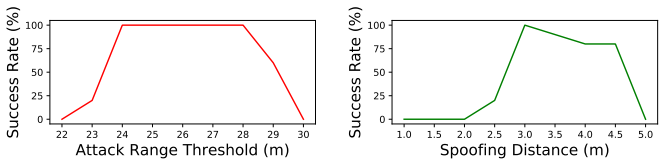

Fig. 4: Attack success rates under different attack range thresholds (left) and spoofing distances (right).

the attacker needs to spoof the GPS such that the ROI covers the back green TL but not the front red TL. Thus, we start by exploring the attack parameters for this scenario. The left figure in Fig. 4 shows the attack success rates when using different attack range thresholds, i.e., the distance from the AV to the TL that triggers the attack. During the experiments, a spoofing distance of 3 meters is applied to the victim's GPS inputs. As shown, the best attack range threshold falls between 24 and 28 meters. When using a small attack range threshold (e.g., 22 meters), the attack fails because the victim is too close to the intersection, and it already executed the stop decision based on the TL detection prior to the attack. On the other hand, if the attack range threshold is too large (e.g., over 28 meters), the victim might already be deviated out of road boundary since the AD system is constantly correcting any deviation between the localization and the planned trajectory [6].

The right figure in Fig. 4 shows the attack success rates when using different spoofing distances. Based on the previous experiments on the attack range threshold, we launch the attack when the victim AV is 26 meters away from the TL. As shown in Fig. 4, a spoofing distance of smaller than or equal to 2 meters is too small to move the red TL out of the ROI, so the victim can still detect it. On the other hand, when the spoofing distance is larger than 4.5 meters, both front and back TLs are moved out of the ROI, causing the victim to fail to detect any TL and simply stop in front of the intersection. When the spoofing distance is 3–4.5 meters, the attack success rate is at least 80%, where the front red TL is moved out of the ROI, but the back green TL is still in and thus detected by the victim AV, causing it to continue driving forward to run the red light. In particular, a 100% attack success rate is achieved when using a spoofing distance of 3 meters.

In the green light detection scenario, instead of controlling the ROI to be at a particular position, the attacker can simply spoof the ROI to the area without any TLs. In such a case, the TL detector will report "unknown" as no TLs exist in the ROI. Consequently, the planning module in the AD system issues a stop decision conservatively. This results in a DoS attack since the victim stops at the green TL. Similarly, we evaluate this scenario and found that our ROI attack can achieve a 100% success rate when using a spoofing distance of 5 meters.

Attack demos. We create two attack demos to illustrate the severe consequences of the ROI attack. Fig. 5 and Fig. 6 show the snapshots when attacking the red and green light detections, respectively. The left sub-figures show the camera images with the ROI annotations, and the right sub-figures show the corresponding driving decisions in Baidu Apollo. As shown, our attack can successfully fool the victim AV to detect a wrong TL or fail to detect any TLs due to the ROI position shifts caused by GPS spoofing. Such an attack poses great dangers to road safety since it can cause the AV to violate traffic rules and may even lead to car crashing consequences, e.g.,