

CSC 144V: Autonomous Vehicles

Lecture 4.2

Adversarial Attacks for AD Exercises Solution

Department of Computer Science,
Hofstra University

Exercise: Tracker and max_age

A Kalman tracker runs at 10 Hz with **max_age = 5**: a track that is unmatched for **more than** 5 consecutive frames is deleted. Between matches the track is predicted forward with a constant-velocity model.

- (a) For how many consecutive frames must the attacker suppress the detections of an object to delete its track? How many seconds is that?
- (b) Suppressing 2 frames: is the track deleted?
- (c) Per-frame suppression succeeds with probability p (independent frames). What is the probability that 6 consecutive frames are all suppressed?

p (per frame)	P(6 frames suppressed)
0.8	
0.9	
0.98	

Exercise Solution: Tracker and max_age

- (a) Deleted when unmatched > 5 frames: **6 frames = 0.6 s** (5 frames are still survived).
- (b) **No**. The filter predicts through the gap and re-matches the next detection; nothing changes.
- (c) $P = p^6$: $p = 0.8 \rightarrow 0.26$; $p = 0.9 \rightarrow 0.53$; $p = 0.98 \rightarrow 0.89$.
- A detection attack that works only 90 % of the time deletes the track just half of the time: attacks must reason about association and track management (Jia et al.: > 98 % detection success may be needed).

p (per frame)	P(6 frames suppressed)
0.8	0.26
0.9	0.53
0.98	0.89

Exercise: LiDAR spoofing delay

A LiDAR measures $R = c \cdot \Delta t / 2$ ($c = 3 \times 10^8$ m/s). An attacker's photodiode detects the outgoing pulse, waits a delay τ , then fires a spoofed return. The attacker is at distance $d = 20$ m from the LiDAR (assume zero electronics delay).

- (a) What round-trip time Δt does the LiDAR measure for the spoofed return, and what apparent range R follows?
- (b) Compute τ so that a phantom appears at 35 m, in the ego lane.
- (c) Can this attacker make a phantom appear at 15 m? Why or why not?

Exercise Solution: LiDAR spoofing delay

- (a) The pulse needs d/c to reach the attacker, then τ , then d/c back: $\Delta t = 2d/c + \tau$, so $R = d + c\tau/2$.
- (b) $35 = 20 + (3 \times 10^8 \cdot \tau)/2 \rightarrow \tau = 15 / (1.5 \times 10^8) = 100 \text{ ns}$.
- (c) **No** (for a reactive attacker): the earliest possible return is $\tau = 0$, i.e. $R = d = 20 \text{ m}$. A closer phantom would need the return to arrive before the pulse reached the attacker.
- The sensor trusts any return with the right wavelength and timing; the delay τ chooses the apparent range.

Exercise: Attack taxonomy

Levels of attack success: **1** model-output error, **2** persistent perception failure (survives tracking and fusion), **3** closed-loop safety violation (collision, lane departure).

For each attack, fill in what the attacker modifies, which stage is attacked, and the highest level shown.

Attack	What the attacker modifies	Stage attacked	Highest level shown
Laser spoofing (CCS 2019)			
Dynamic patch vs. MOT			
Dirty road patch			
Acero			

Exercise Solution: Attack taxonomy

Levels: **1** model-output error, **2** persistent perception failure, **3** closed-loop safety violation.

Dirty road patch: > 97.5 % success, 100 % collision rate in software-in-the-loop and on a real vehicle. Acero: 73.3 % of attacks end in a collision (simulation). **Adversarial training and sensor checks are irrelevant to Acero**: no pixel, point or signal is modified.

Attack	What the attacker modifies	Stage attacked	Highest level shown
Laser spoofing (CCS 2019)	Injected LiDAR points	LiDAR detection	Level 1
Dynamic patch vs. MOT	Pixels (adversarial patch)	Detection + tracking	Level 1-2
Dirty road patch	Road appearance (physical patch)	Lane detection → steering	Level 3
Acero	Nothing: normal-looking driving maneuvers	Planning / control of the victim	Level 3

Exercise: Lane-centering toy model

Toy model (Dirty Road Patch): speed **15 m/s**, lane half-width **1.8 m**, car half-width **0.95 m**. A road patch biases the perceived lane center by b ; the controller faithfully tracks the perceived center. Assume the car drifts laterally toward the wrong center at **0.3 m/s**.

- (a) What is the lateral margin before a wheel crosses the lane boundary?
- (b) After how many seconds does the car cross the boundary, and how far has it travelled?
- (c) Does the car leave the lane for $b = 1.4$ m? For $b = 0.5$ m? (The car ends up offset by b .)

Exercise Solution: Lane-centering toy model

- (a) Margin = $1.8 - 0.95 = 0.85$ m.
- (b) $t = 0.85 / 0.3 \approx 2.8$ s; distance = $15 \times 2.83 \approx 42.5$ m.
- (c) $b = 1.4$ m > 0.85 m: **leaves the lane**. $b = 0.5$ m < 0.85 m: **stays** inside (margin only 0.35 m left).
- The attack objective must be defined on the **control error** (does the car cross the line?), and evaluated with a vehicle model in the loop, not on the lane-detection error alone.