Securing Autonomy
Study of Perception-Based Control

Miroslav Pajic

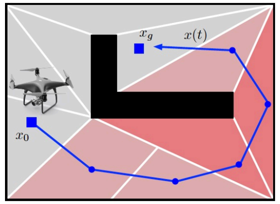
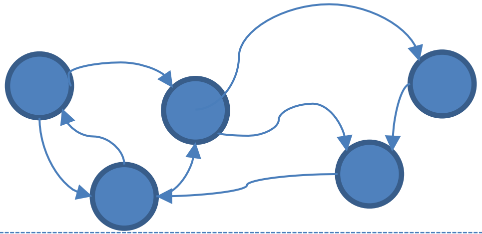
CPSL@Duke
Department of Electrical and Computer Engineering
Department of Computer Science
Duke University

Duke
PRATT SCHOOL of ENGINEERING

DUKE ROBOTICS

Security-Aware Autonomy

Duke PRATT SCHOOL of ENGINEERING

Control Stack	Control view	Modeling view
Mission Planner	Long-horizon views	
Tactical Planner	Short-horizon views	
Low-level Control	Continuous/discrete control with constraints	$f_r(x(t)) = \int_0^T \varrho(x(t), h(t))dt + \int_0^T \lVert x(t) \rVert^2 dt,$ $\min \; f_r(x_r(t)) + f_h(x_h(t))$ $\text{s.t.} \; x_r(t) = x_h(t), \; u_r(t) = u_h(t),$
Vehicle		

Adding Resiliency

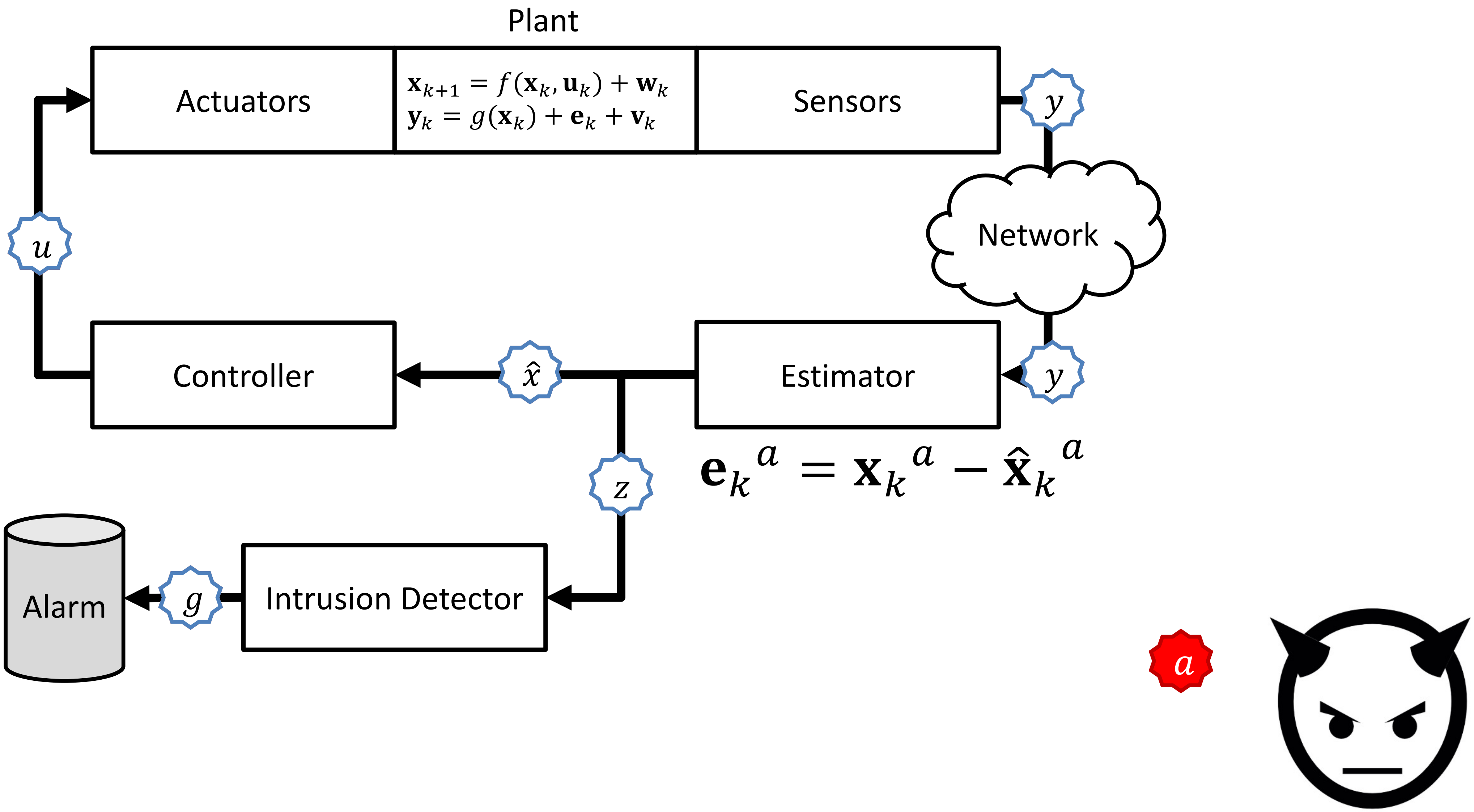
[USENIX Sec'22*, ICCPS22b*, CDC21, ICRA21a, ICRA21b, ICRA20, ICRA19, CAV'19a, THMS19]

[Automatica21*, TII21, TASE21, CDC19a, CDC19b, IoTDI19]

[ICCPS22a*, TCPS20, ACC20, AUT21b, AUT21a, AUT18, TECS17, RTSS17, TCNS17a, TCNS17b, CSM17, CDC17, CDC18,…]

Our Goal: Add resiliency to controls across different/all levels of the autonomy stack

Low-Level Control in the Presence of Attacks



Can Attacker Reach Any State?

Duke PRATT SCHOOL of ENGINEERING

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k$$
$$\mathbf{y}_k = C\mathbf{x}_k + \mathbf{a}_k + \mathbf{v}_k$$

$$supp(\mathbf{a}_k) = \mathcal{K}$$
$$\mathbf{a}_{k,i} = 0, \forall i \in \mathcal{K}^C$$

Theorem 1 [1,2,3,**4,5**]:
A system presented above is perfectly attackable if and only if it is **unstable**, and at least one eigenvector **v** corresponding to an unstable mode satisfies $supp(\mathbf{Cv}) \subseteq \mathcal{K}$ and **v** is a reachable state of the dynamic system.

Physics-based detectors cannot always protect us from an intelligent attacker

[1] Y. Mo and B. Sinopoli, "*False data injection attacks in control systems*," in First Workshop on Secure Control Systems, 2010
[2] C. Kwon, W. Liu, and I. Hwang, "*Analysis and design of stealthy cyber attacks on unmanned aerial systems*", J. of Aerospace Inf. Systems, 2014
[3] I. Jovanov and M. Pajic, "*Relaxing Integrity Requirements for Attack-Resilient Cyber-Physical Systems*", IEEE Trans. on Automatic Control, 2019
[4] A. Khazraei and M. Pajic, "*Perfect Attackability of Linear Dynamical Systems with Bounded Noise,*" ACC 2020.
[5] A. Khazraei and M. Pajic, "*Attack-Resilient State Estimation with Intermittent Data Authentication,*" Automatica, 2021.

What happens when we include perception?

Vulnerability of Perception

- ***Deep Learning is workhorse in modern perception pipelines***
- ***Attacks on perception studied at single sensor, single time-instance level; LiDAR underrepresented***

  ↳ *Not representative of real systems or adv. objectives!*

  - *Real systems use sensor fusion across multiple sensors and multiple time points; rely heavily on LiDAR*
  - *Adv. Objectives include creating false objects, removing existing objects, or translating existing objects --> very few systematic evaluations of all outcomes*
- *Sensor fusion claimed to be "resilient", often "silver-bullet" for defense but this claim rarely experimentally validated*

Point cloud (LiDAR) data & algorithms are under-analyzed in the security community

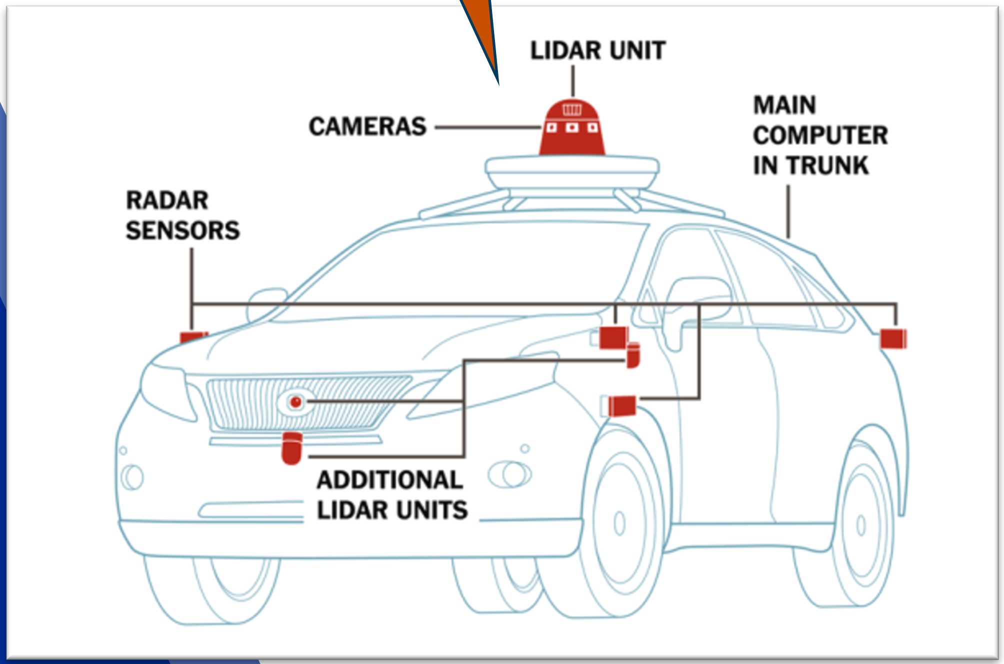
Sensor fusion (e.g. fusion at data-level, tracking-level) must be analyzed due to ubiquitous adoption across industry

Saturation Attacks

Clear-Box Adversarial Perturbation

Clear-Box Adversarial Injections

Structured Injection and Spoofing



Camera-LiDAR Fusion

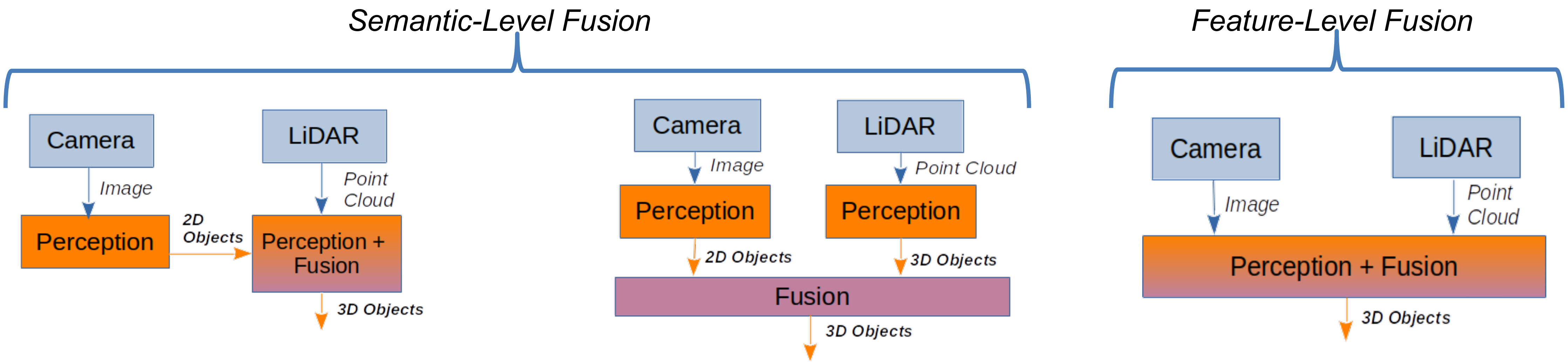
Multiple Architectures for Sensor Fusion

Duke
PRATT SCHOOL *of* ENGINEERING

- *Semantic fusion* popular across industry due to:
  - Reduce of "curse of dimensionality" of input space
  - Greater flexibility in industry for "plug-and-play"/swap-ability of components
- *Feature-level-fusion* high-performing due to fusion of low-level, machine-learned features
- Fusion touted to improve resiliency and performance compared to single-sensor perception alone

Most common sensors:

- LiDAR data is sparse in R4
  - X-Y-Z-intensity
  - Full 3D resolution
- Camera data is dense in R3
  - R-G-B channels
  - 2D (angles-only) resolution



Cascaded semantic fusion

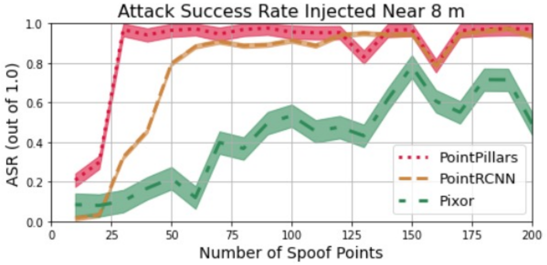
Fusion at tracking
"Integrated semantic fusion"

Feature-level-fusion

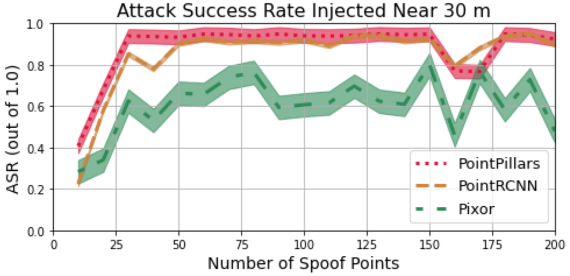
Find Fusion On-Par With Existing Defenses Against Naïve Spoofing Attacks

Duke
PRATT SCHOOL of ENGINEERING

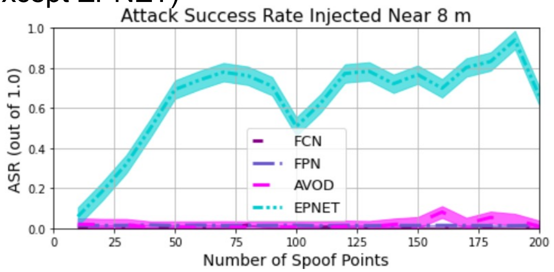
Undefended attack success high against LiDAR-only percep at close range



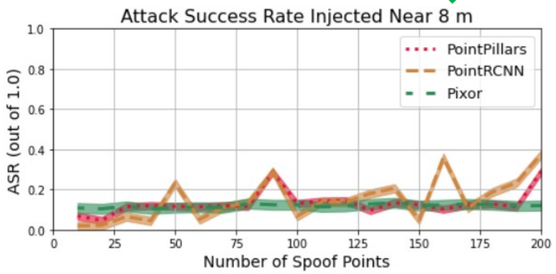
***Undefended attack success high against LiDAR-only percep at med. range**



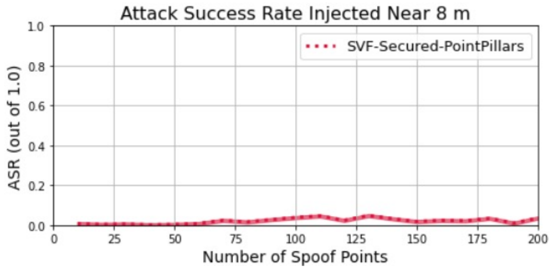
***Fusion guards against naïve attack at close range** ✓
(except EPNET)



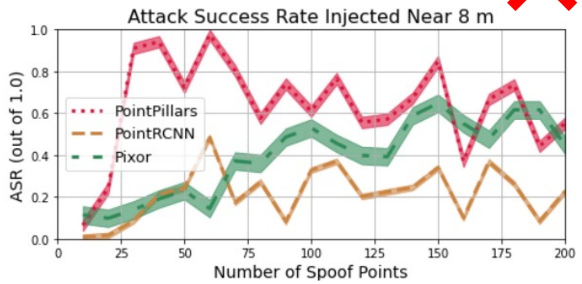
CARLO guards against naive attack at close range ✓



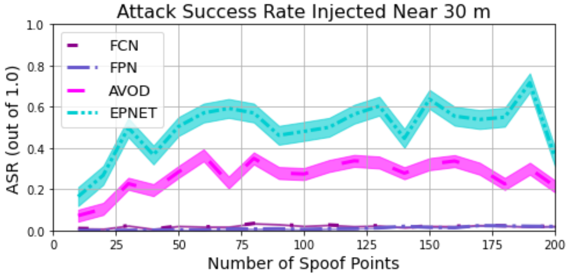
SVF guards against naïve attack at close range ✓



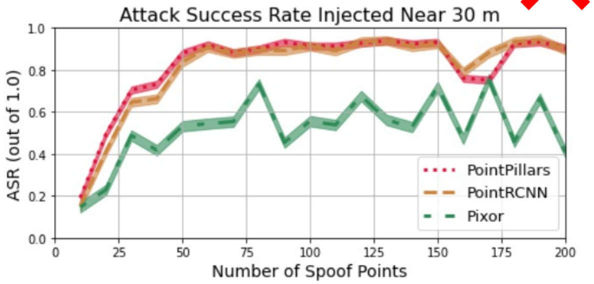
***ShadowCatcher does not guard against naïve attack at close range; has high induced FN rate (not shown)** ✗



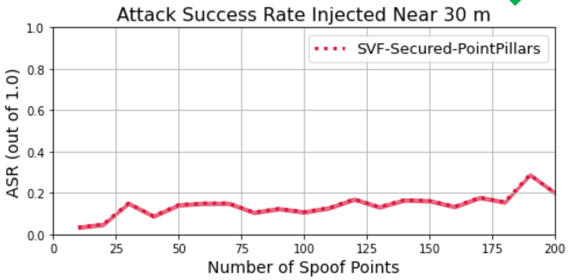
***Fusion guards against naïve atack at med. Range** ✓
(except EPNET; AVOD performs ok)



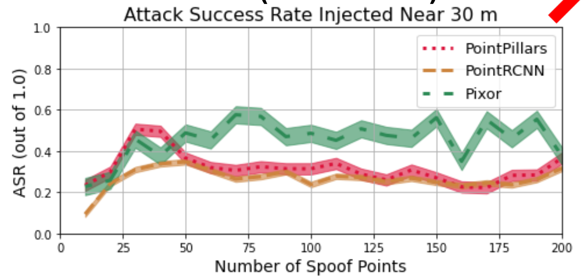
***CARLO does not guard against naïve attack at medium range** ✗



***SVF guards against naïve attack at medium range** ✓



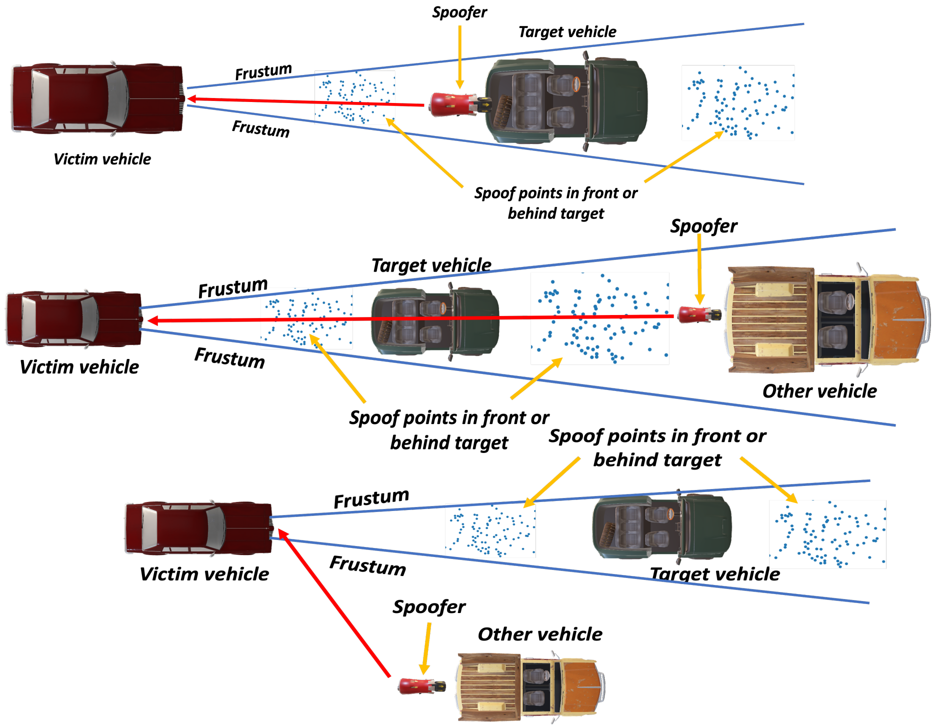
***ShadowCatcher does not guard against naïve attack at med. range; has high induced FN rate (not shown)** ✗



**Novel contribution of our work*

Beyond Naïve Attack: Novel Frustum Attack Is Feasible

Three candidate realizations of the frustum attack.
Additional configurations shown later



Compromise Fusion (and LiDAR-only)

- Fusion robust against naïve attack because naïve attack is not consistent between sensor modalities
- Ensure consistency by spoofing *within the frustum (i.e. in-view, as seen by camera)* of existing vehicles
- This does not require any knowledge of the camera data

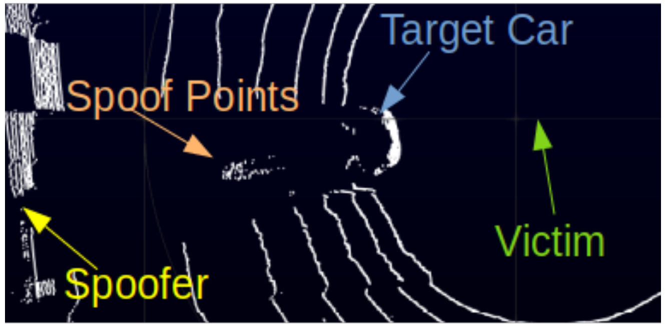
Feasibility

- We validated attack feasibility with limited additional knowledge required over original, naïve black-box spoofing
- Only additional requirement is attack orientation

Target car in front of victim

Spoofer set behind target car

Stable spoof points placed in frustum



Demonstrated controlling (i.e. moving to attacker's specified location) spoof points stably over time with moving vehicles

Frustum Attack is Widely Successful

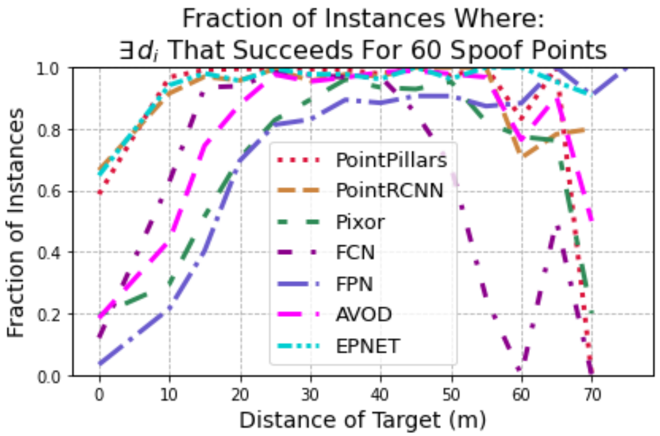
Duke
PRATT SCHOOL of ENGINEERING

Compromise Fusion (and LiDAR-only)

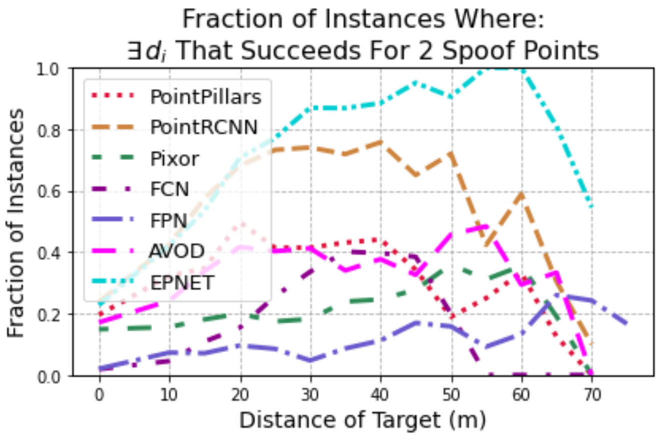
- Frustum attack demonstrated to compromise BOTH LiDAR-only AND camera-LiDAR fusion
- Frustum attack shown indefensible by state-of-the-art defenses (CARLO, SVF, ShadowCatcher, LIFE)

Extensive Evaluations

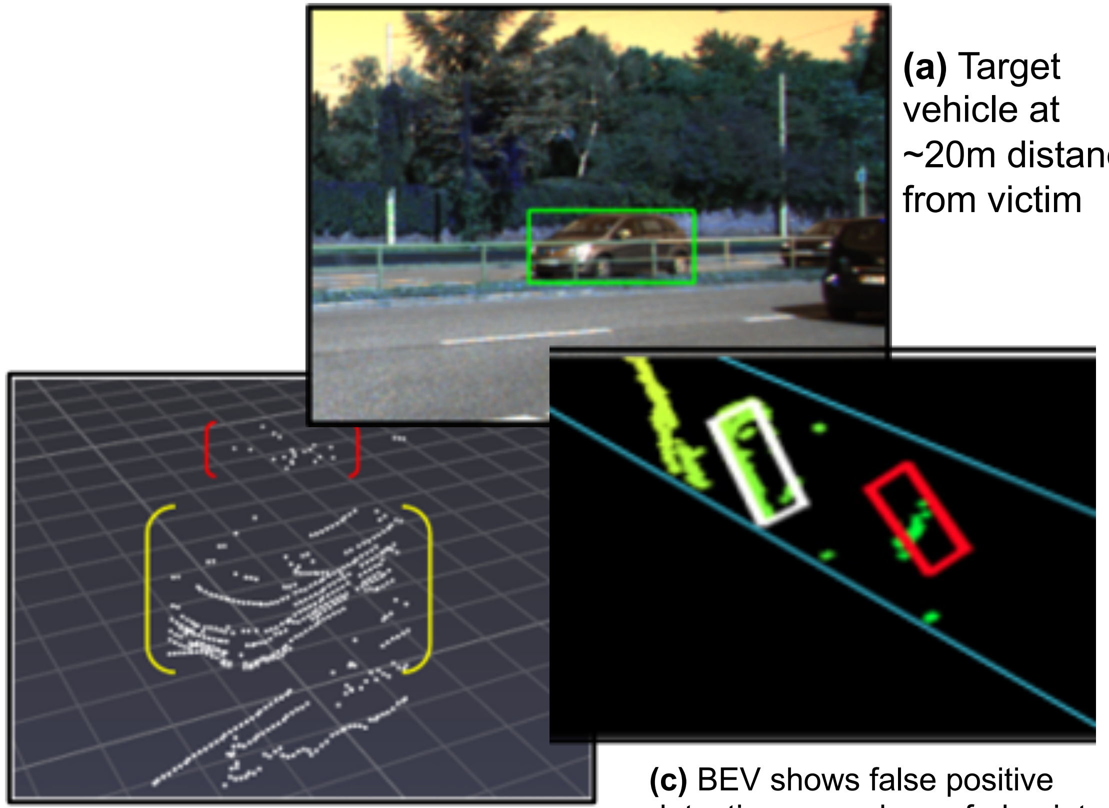
- We perform the most extensive evaluation of attacks on perception to-date with 8 algorithms and 4 defenses (7 and 3 for large-scale evaluation)
- > 75 million attack traces evaluated --> number of spoof points, distance of spoof point placement, each object, each frame of data



Frustum attack widely successful with 60 spoof points



Frustum attack successful even with just 2 spoof points!



(a) Target vehicle at ~20m distance from victim

(b) Target victim (yellow, 238 pts) has many more points than the spoof points (red 20pts)

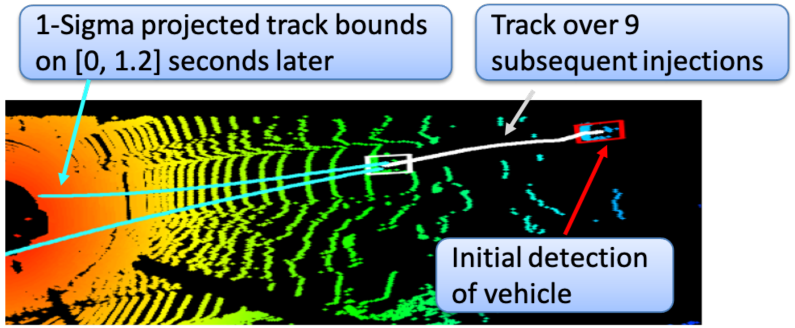
(c) BEV shows false positive detection around spoofed points

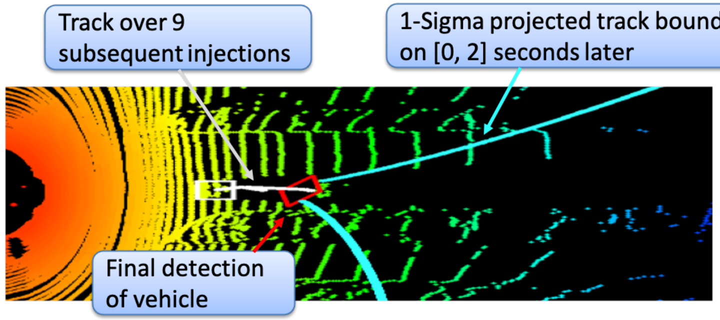
Longitudinal Frustum Attacks Are Dangerous

Duke
PRATT SCHOOL of ENGINEERING

Evaluation of Multi-Frame Tracking

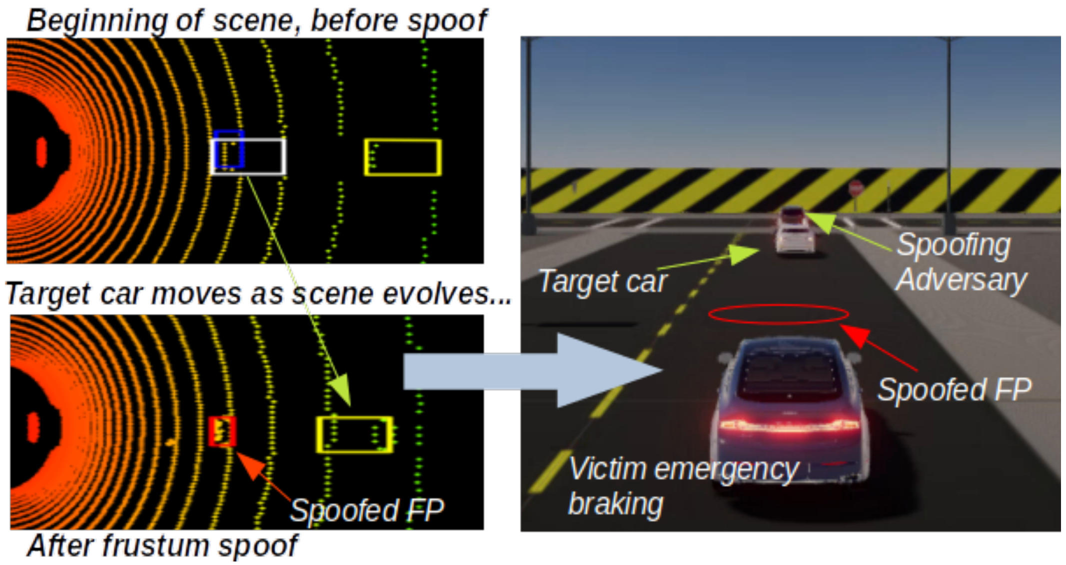
- Use captured KITTI dataset to evaluate impact of frustum attack over multiple frames
- Demonstrated stably executing frustum attack in longitudinally-consistent way to obtain adversarial tracks (white + cyan) that can:
  - 1) project to collide with victim
  - 2) project to accelerate flow of traffic





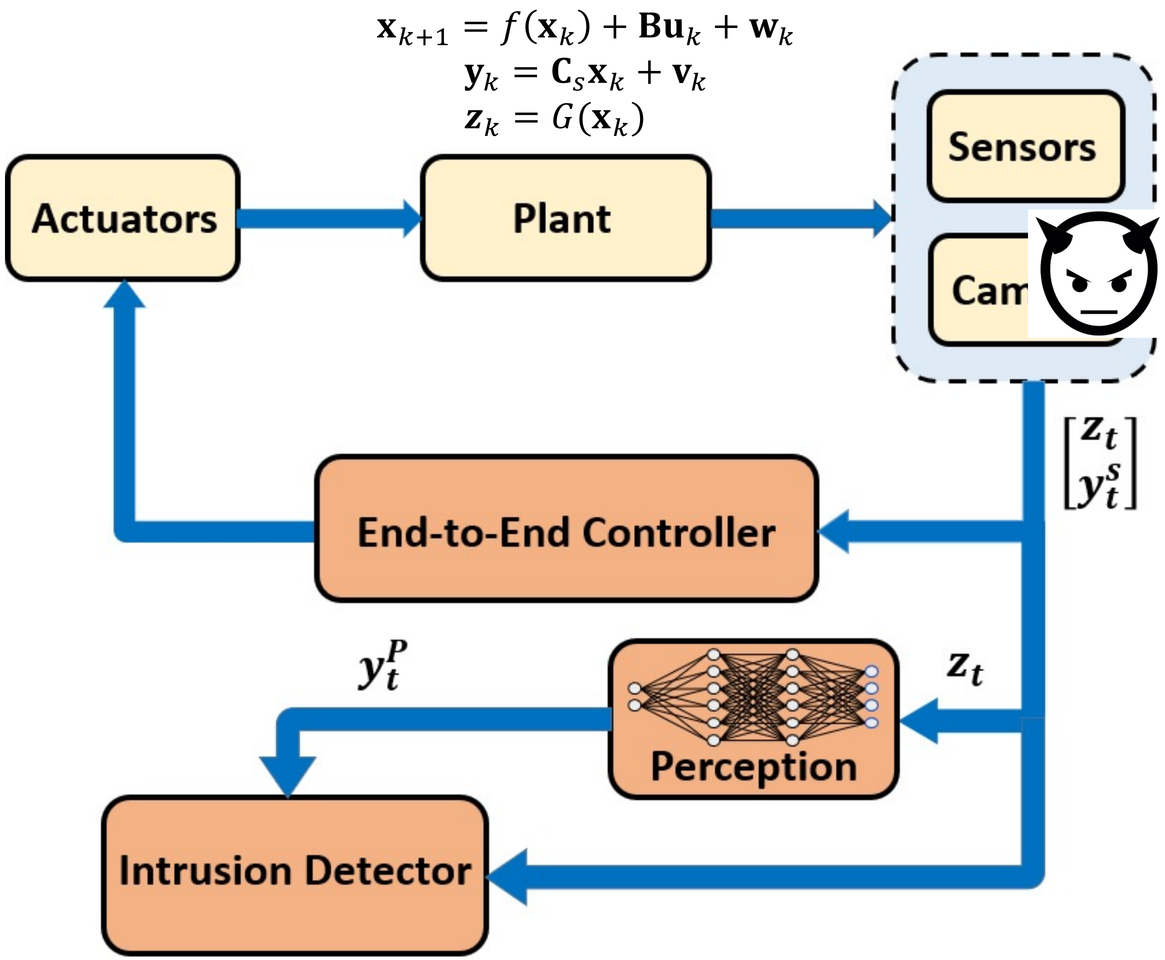
End-to-End, Industry-Grade AVs

- Preliminary evaluation of the vulnerability of Baidu Apollo perception + control stack to the frustum attack – *emergency braking engaged*
  - Baidu fuses LiDAR and camera detections at the tracking-level
  - Use multi-stage approach since Baidu+SVL combination is still under development
- Physics-based simulations of AV driving with the SVL Simulator



Stealthy Spoofing Frustum-Attacks:
Attacking Baidu’s Apollo

So, what happens when we include perception?



Thank you

Duke UNIVERSITY | PRATT SCHOOL *of* ENGINEERING