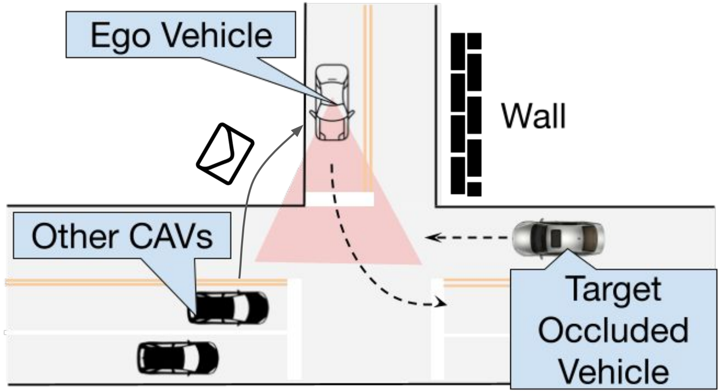
On Data Fabrication in Collaborative Vehicular Perception: Attacks and Countermeasures

Qingzhao Zhang[1], Shuowei Jin[1], Ruiyang Zhu[1], Jiachen Sun[1], Xumiao Zhang[1], Qi Alfred Chen[2], Z. Morley Mao[1]

University of Michigan[1], UC Irvine[2]

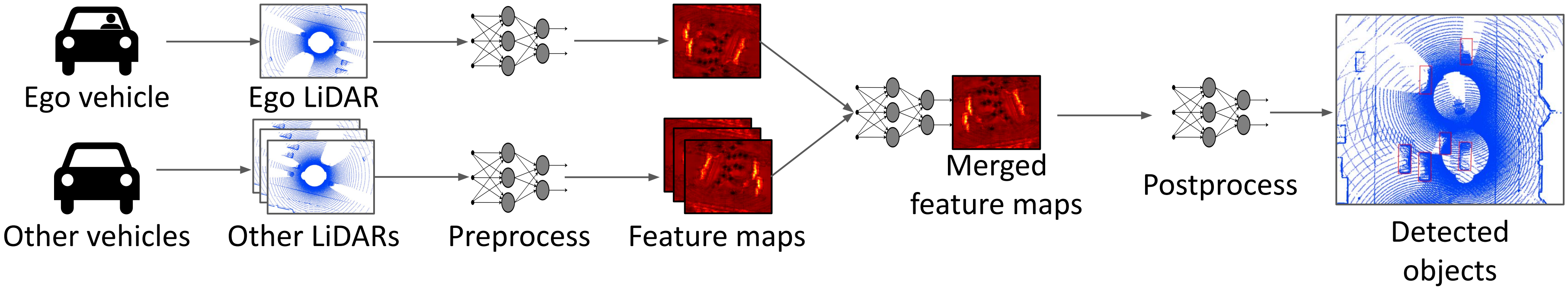
Background: collaborative perception

- Connected and autonomous vehicles share (processed) sensor data to do perception jointly, which enhances perception capability.
  - We focus on Vehicle-to-Vehicle (V2V) sharing of LiDAR data.



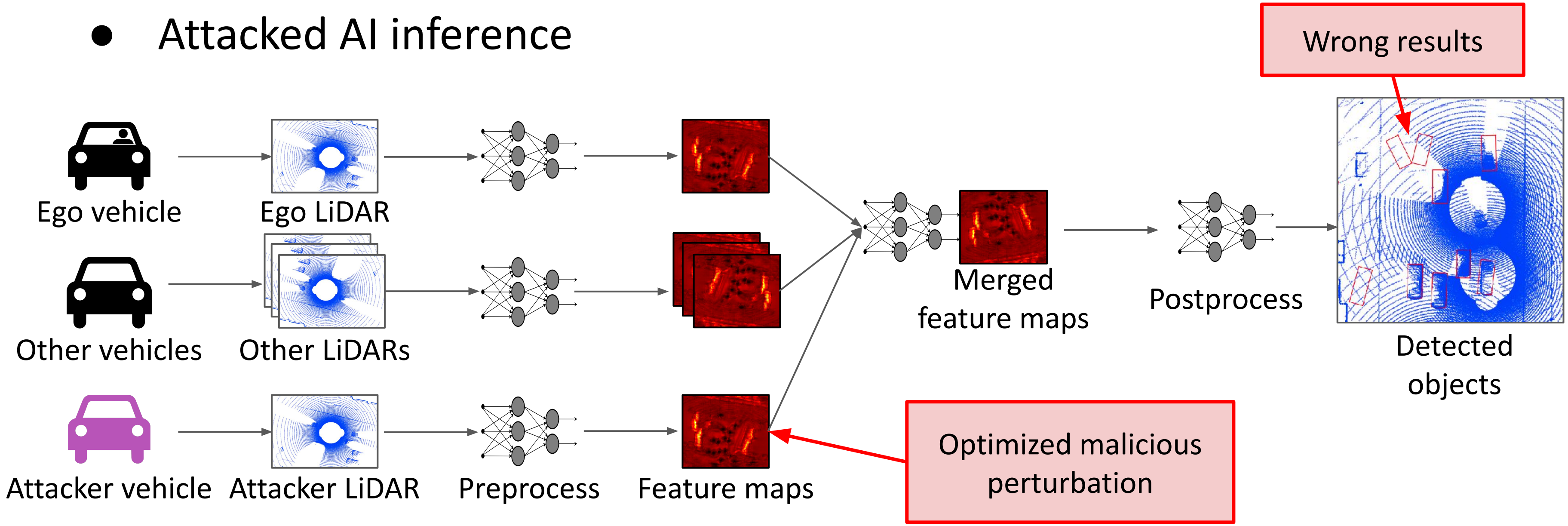
Background: the normal workflow of collaborative perception

- Normal AI inference in each LiDAR cycle



Prior AI adversarial attack

- Attacked AI inference



Tu, James, et al. "Adversarial attacks on multi-agent communication." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Our design of new attacks and countermeasures

- A new attack
  - New **realistic stealthy** attacks to spoof/remove objects at a selected location in collaborative perception

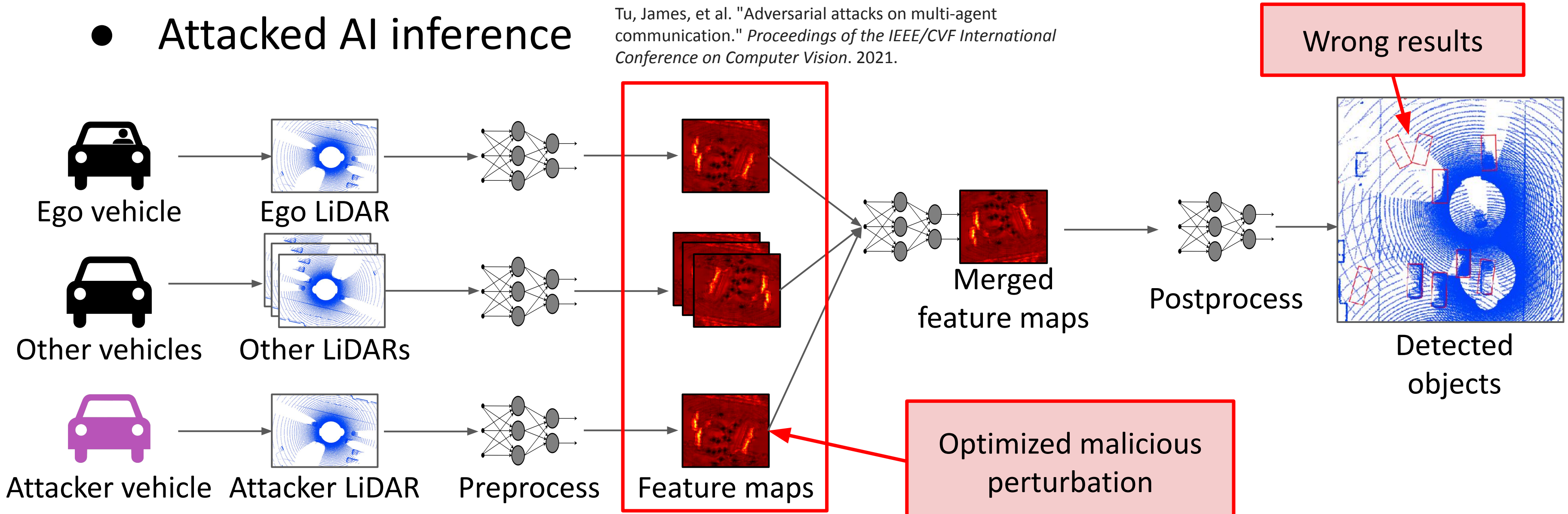
- An anomaly detection method
  - The anomaly detection leverages the collaboration of multiple vehicles to combat against the new threat.

- Our experiments cover both simulation and real-world scenes.

Prior AI adversarial attack is unrealistic

- Attacked AI inference

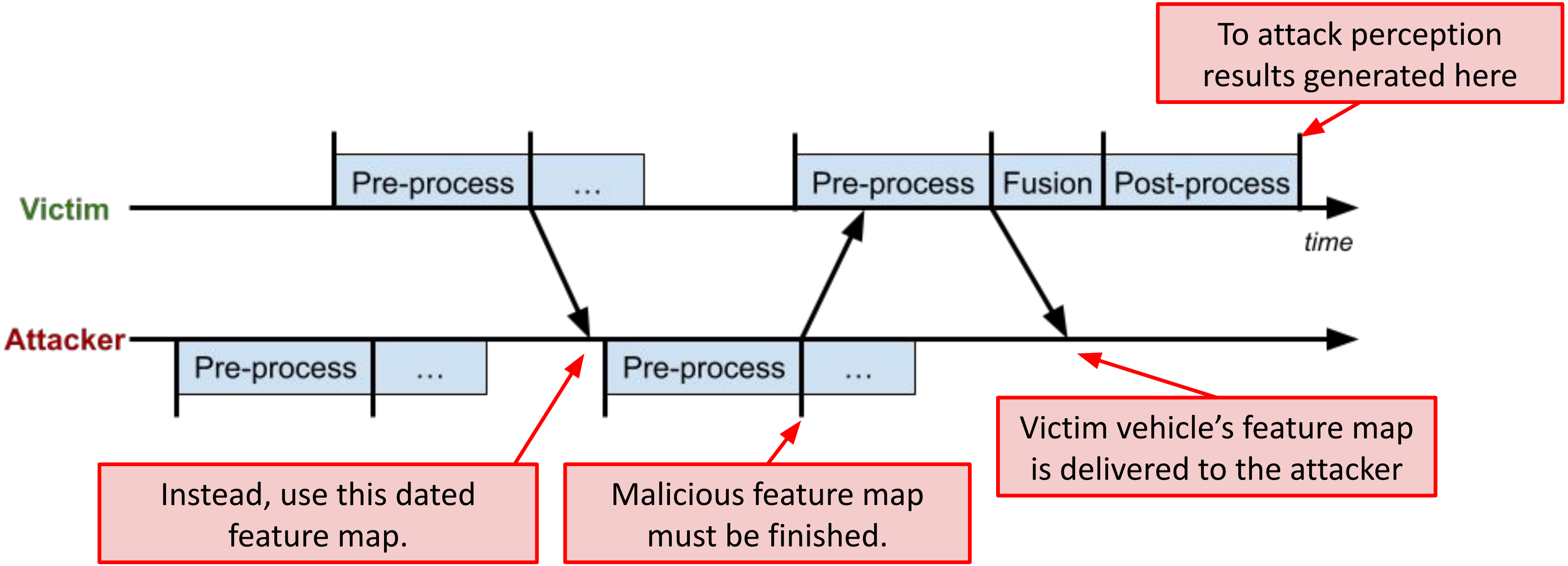
Tu, James, et al. "Adversarial attacks on multi-agent communication." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.



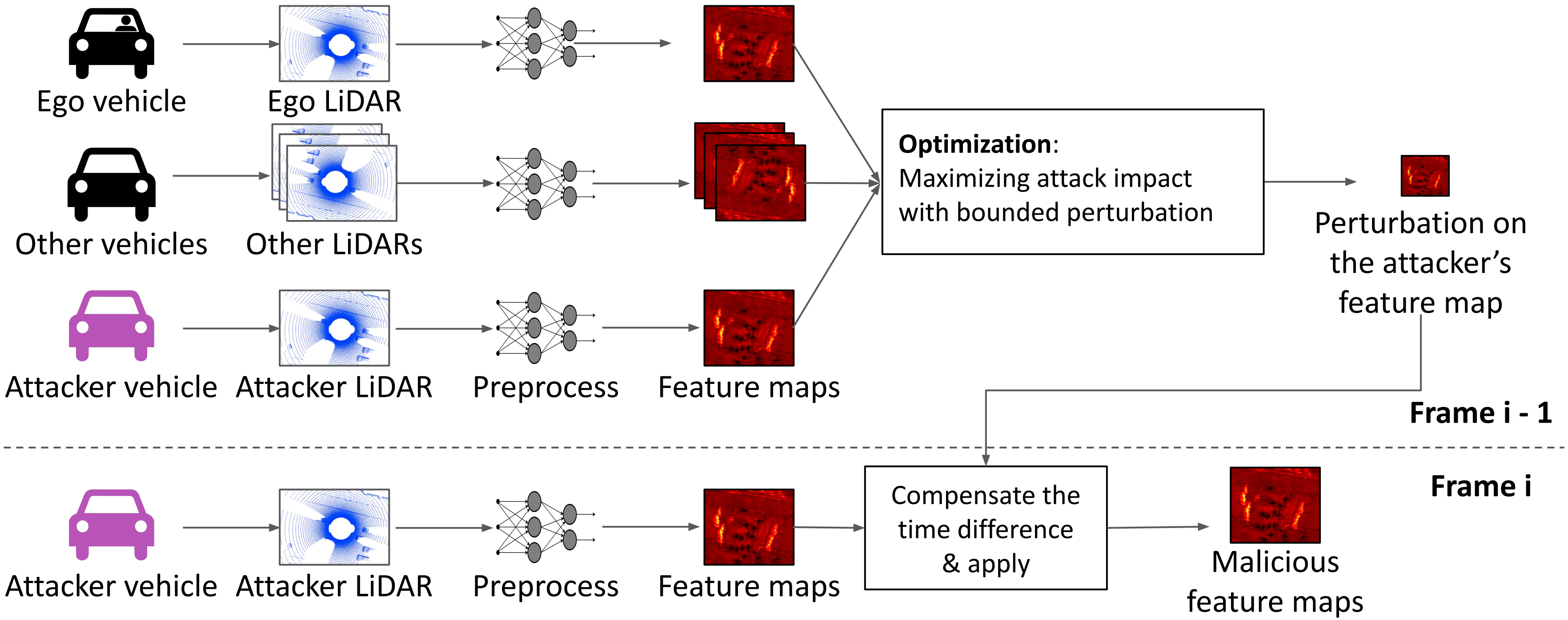
The attacker needs **all feature maps** and the **models** to optimize the perturbation

Prior AI adversarial attack is unrealistic

- Need to consider data transmission latencies and temporal ordering of events.

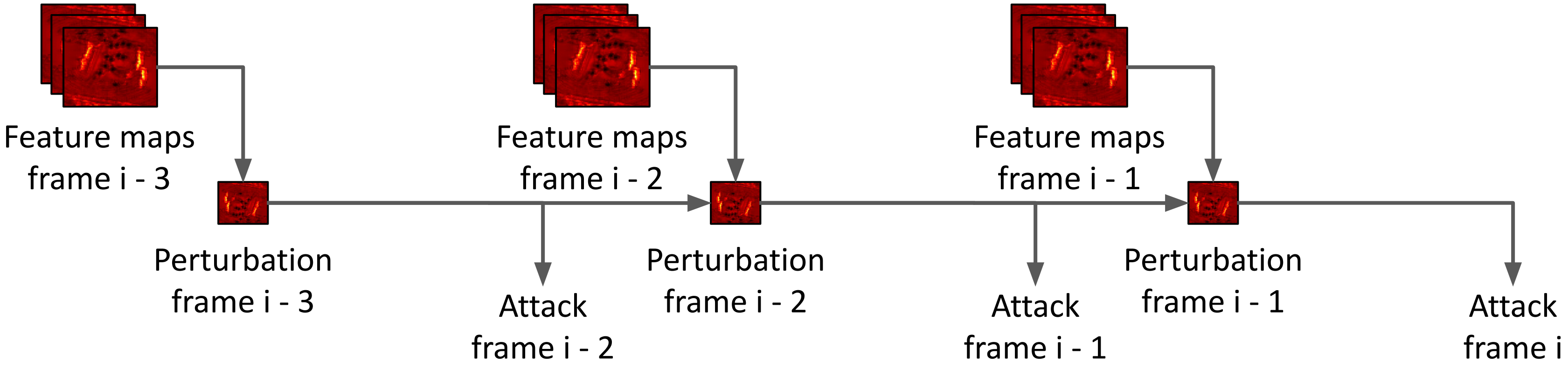


Data flow of our proposed attack scheduling



Reuse optimization results in consecutive frames for efficiency

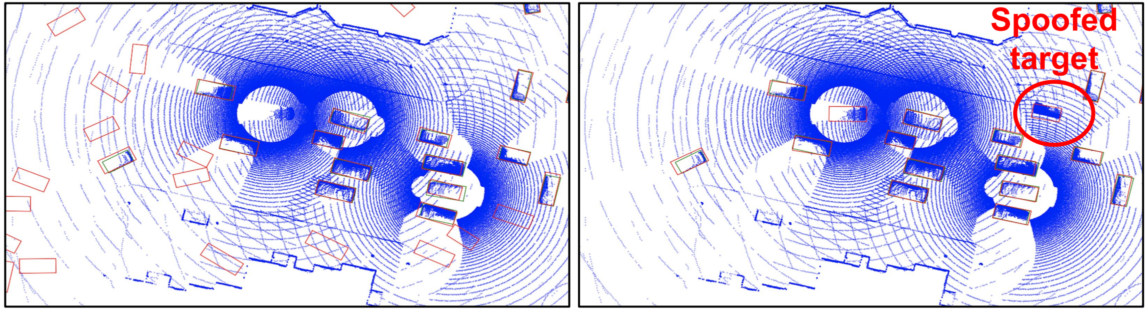
- Strong optimization requires multiple iterations which is still hard to complete in one cycle time (100 ms).
- We can use the optimization results from the last frame to initialize new optimization. One step of optimization for each frame.



Optimization problem for a stealthy targeted attack

- Optimizing a perturbation on the attacker's feature map.
- Maximizing attack impact (spoof or remove an object) in perception results **in a specific targeted region**

Loss function: $$l(Z, z_t) = \sum_{z \in Z} \underbrace{IoU(z, z_t)}_{\text{Box overlap with a target}} \cdot \underbrace{\log(1 - z_\sigma)}_{\text{Confidence score}}$$

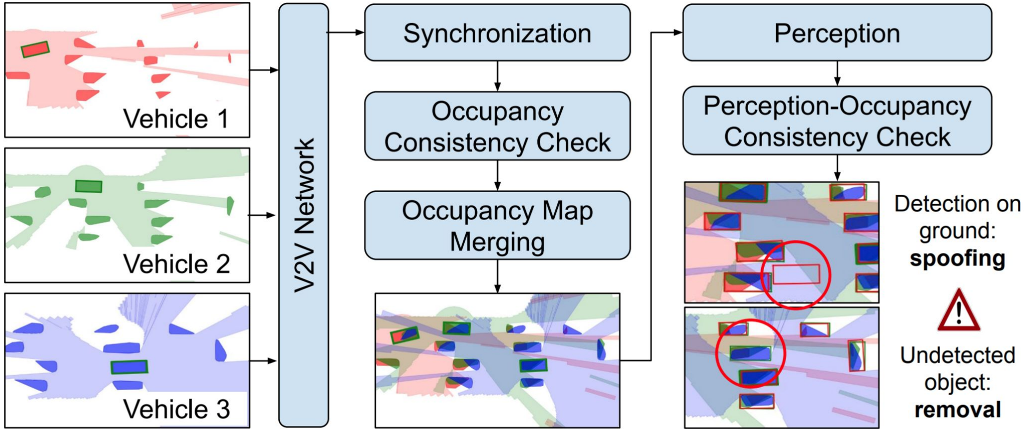


Untargeted adversarial attack (Tu et. al.) | Our targeted adversarial spoofing attack

Tu, James, et al. "Adversarial attacks on multi-agent communication." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Anomaly detection as a mitigation to data fabrication attacks

- Attacked perception results have conflicts with the knowledge of benign CAVs.
- Using occupancy maps to reveal spatial conflicts
  - Detected object on free areas? ⇒ potential spoofing
  - No detected object on an occupied area? ⇒ potential removal



Evaluation on simulation dataset

- The evaluation is on 300 randomly selected attack scenarios from OPV2V dataset [1]

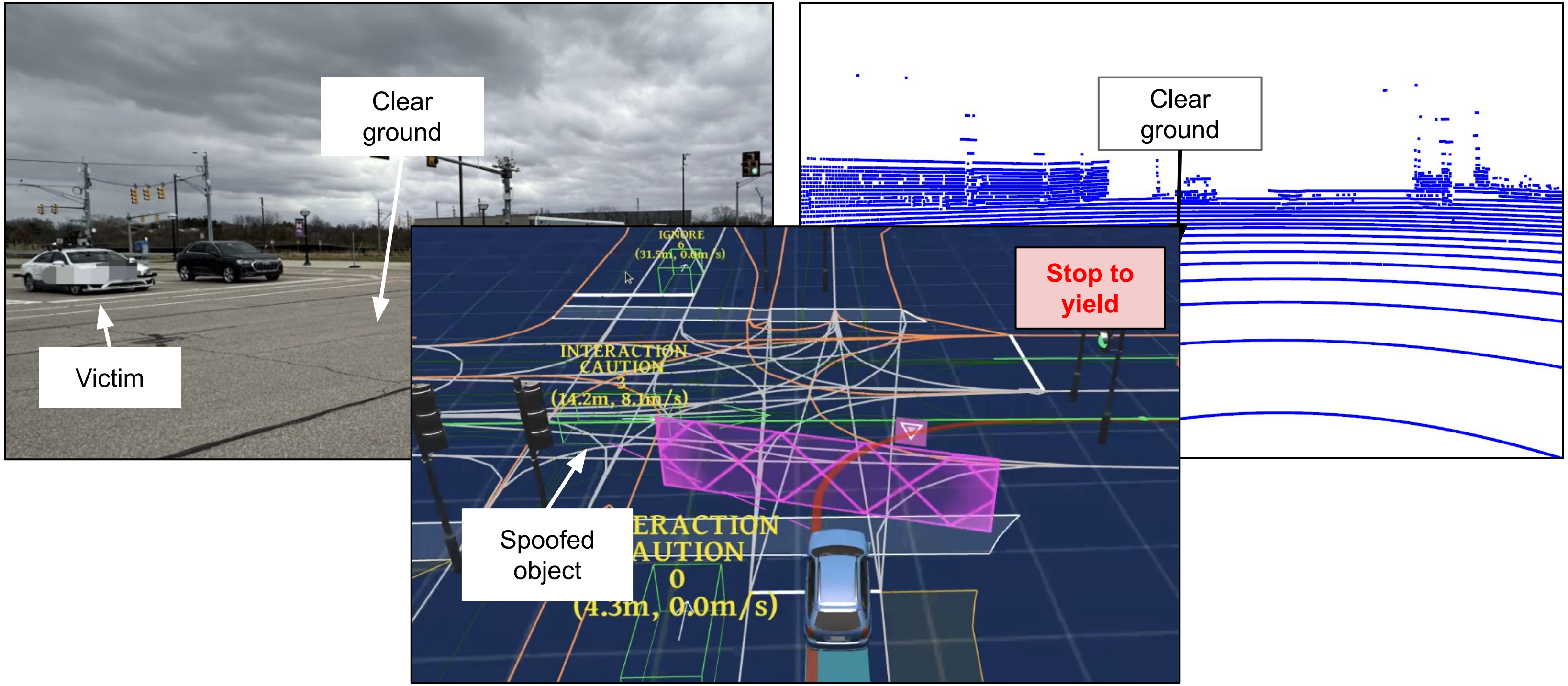
Attack setting: Method-Fusion-Goal	Attack results				Defense results		
	Succ.	IoU	Score	ΔAP	Succ.	TPR	FPR
RC-Early-Spoof	86.0%	0.55	0.38	-0.4%	83.8%	80.9%	2.0%
RC-Early-Remove	87.3%	0.07	0.03	-0.5%	81.2%	38.0%	5.6%
Adv.-Int.-Spoof	90.0%	0.46	0.71	-2.0%	83.4%	80.1%	2.0%
Adv.-Int.-Remove	99.3%	0.02	0.01	-3.9%	83.6%	42.5%	2.2%
Naive-Late-Spoof	98.7%	0.96	0.99	0	80.8%	84.8%	2.7%
Naive-Late-Remove	0.3%	0.78	0.53	0	-	-	-

Notes: *Int.* - intermediate-fusion. *RC* - ray casting. *Adv.* - adversarial attack. *Succ.* - success rate.

[1] Xu, Runsheng, et al. "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication." 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022.

Real-world experiment in MCity testbed

Real-world experiment in MCity testbed



Conclusions

- Realizability of attacks on autonomous vehicles is greatly affected by temporal and spatial constraints of real systems.
- It is a severe vulnerability for vehicles to depend critical perception on untrusted data.
- Future effort in improving security and reliability of collaborative perception is required.

Artifact: https://github.com/zqzqz/AdvCollaborativePerception

EMail: qzzhang@umich.edu

Thank you!