Vehicle Computing: Vision and Challenges

Weisong Shi

Wayne State University
&
University of Delaware

weisong@udel.edu
http://thecarlab.org

Roadmap

☞ Why Vehicle Computing?

- Research activities @CAR Lab

The Era of CAVs

- **CV Market:**
  - **$65 billion** in 2021, **$225 billion** by 2027 with a CAGR of **17%**
  - **Every new** vehicle will be connected by 2025 (**400 million**)
  - **50%** of national vehicles with connected features

Ford and Google to accelerate auto innovation and reinvent connected vehicle experience
March 31, 2021

Honda Partners with Google for In-Vehicle Services
September 28, 2021

Cruise and GM Team Up with Microsoft to Commercialize Self-Driving Vehicles
2021-01-19

Volkswagen Group teams up with Microsoft to accelerate the development of automated driving
Feb 11, 2021

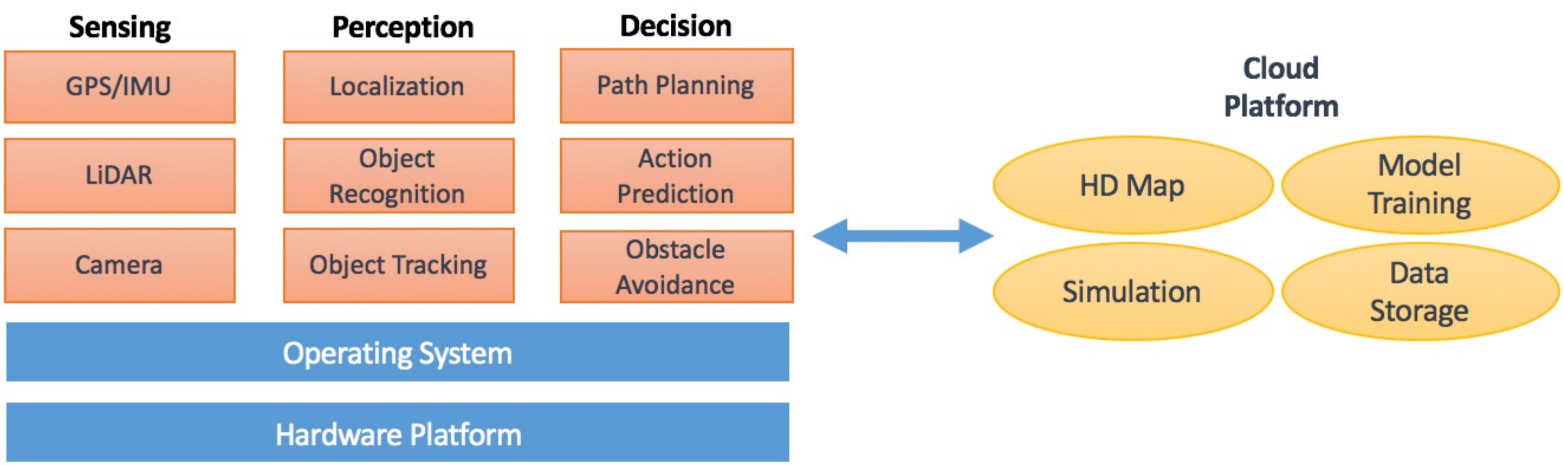
Toyota and Amazon Web Services Collaborate on Toyota's Mobility Services Platform
August 17, 2020

Alibaba launches electric car in tie-up with SAIC
January 13, 2021

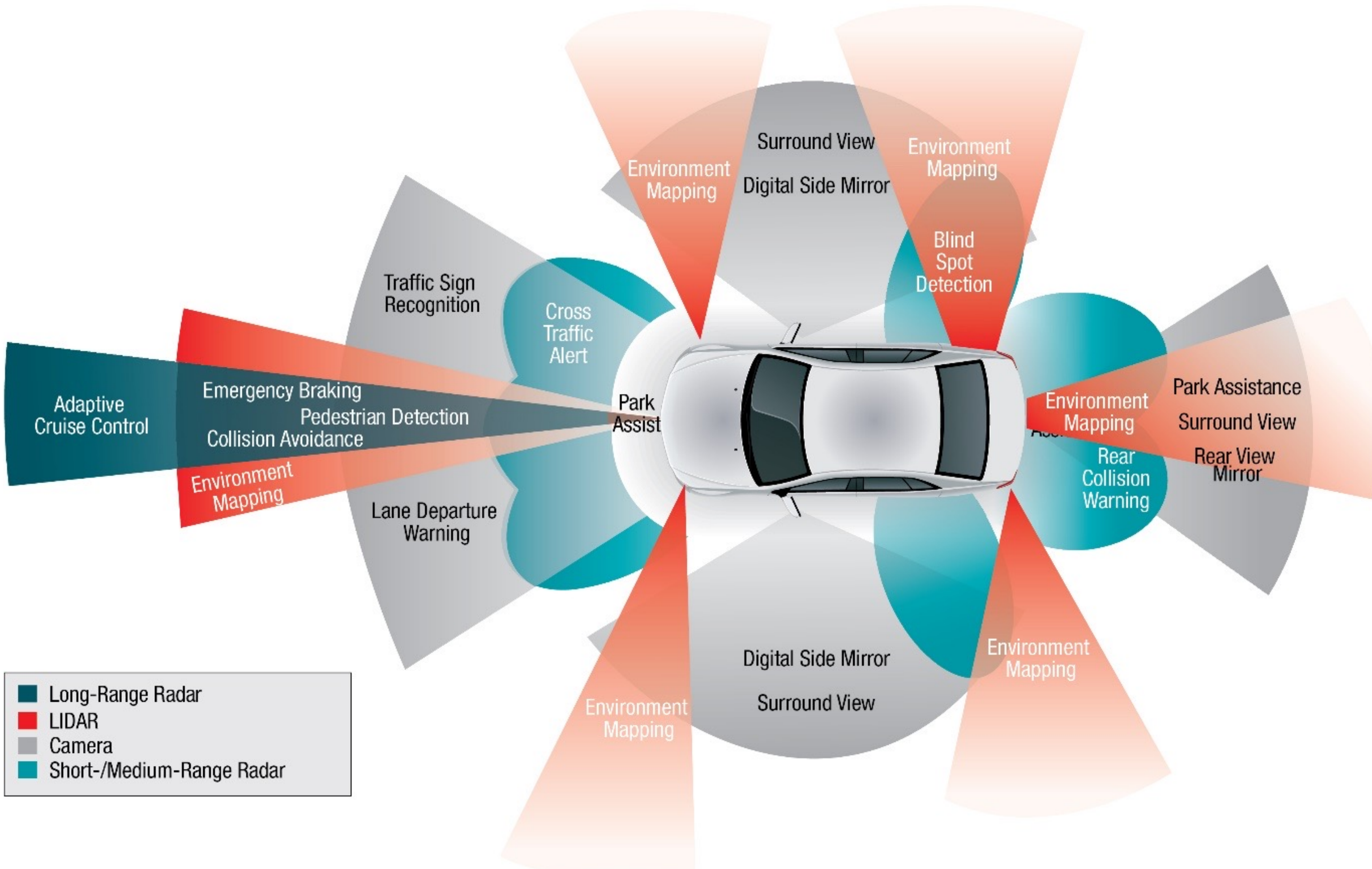
AWS and BMW Group Team Up to Accelerate Data-Driven Innovation
08.12.2020 PRESS RELEASE ARCHIVE

China's Baidu to create an intelligent EV company with automaker Geely
JANUARY 10, 2021

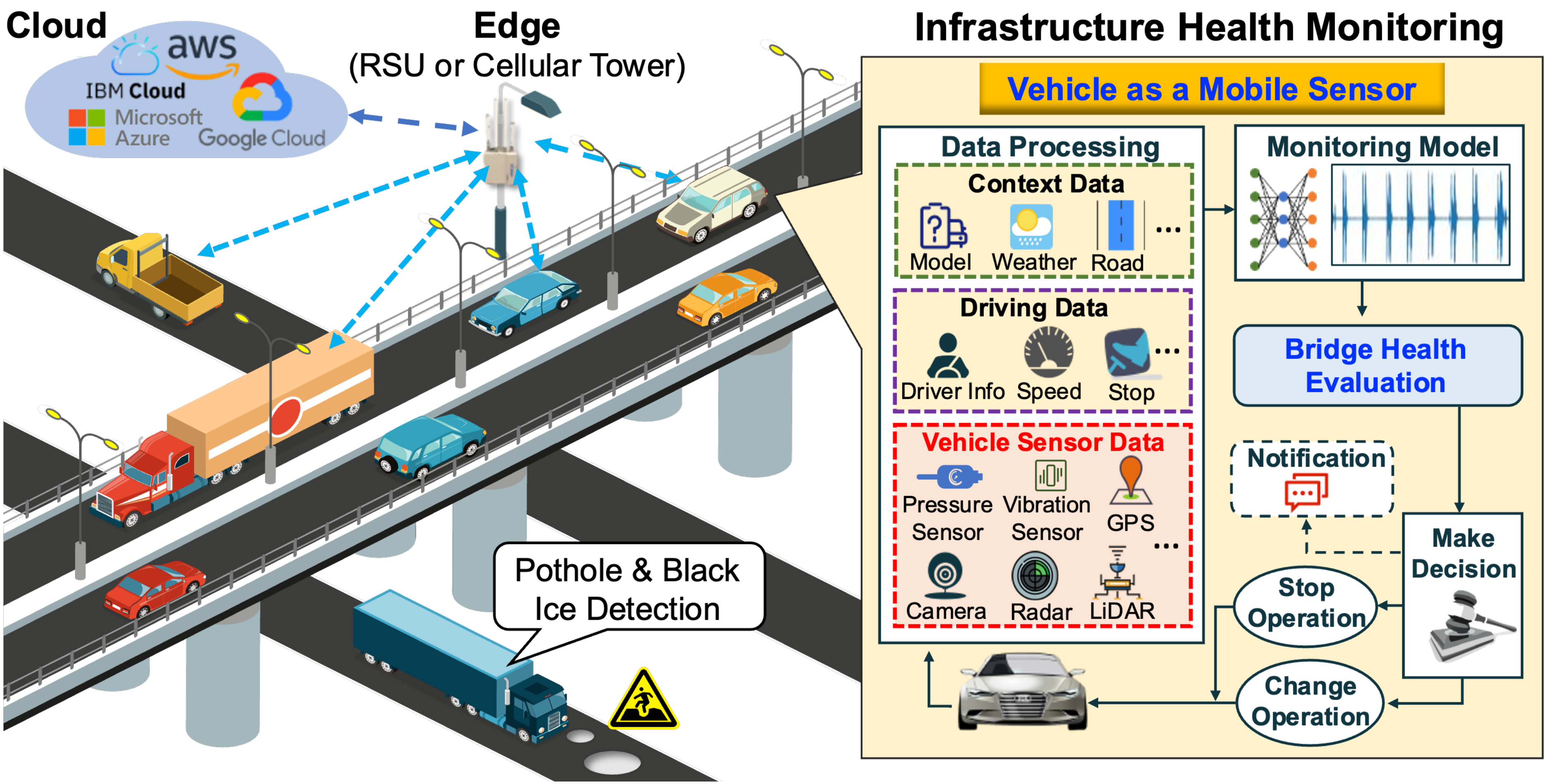
CAV: An Overview



Perception Area of CAVs



Infrastructure Management

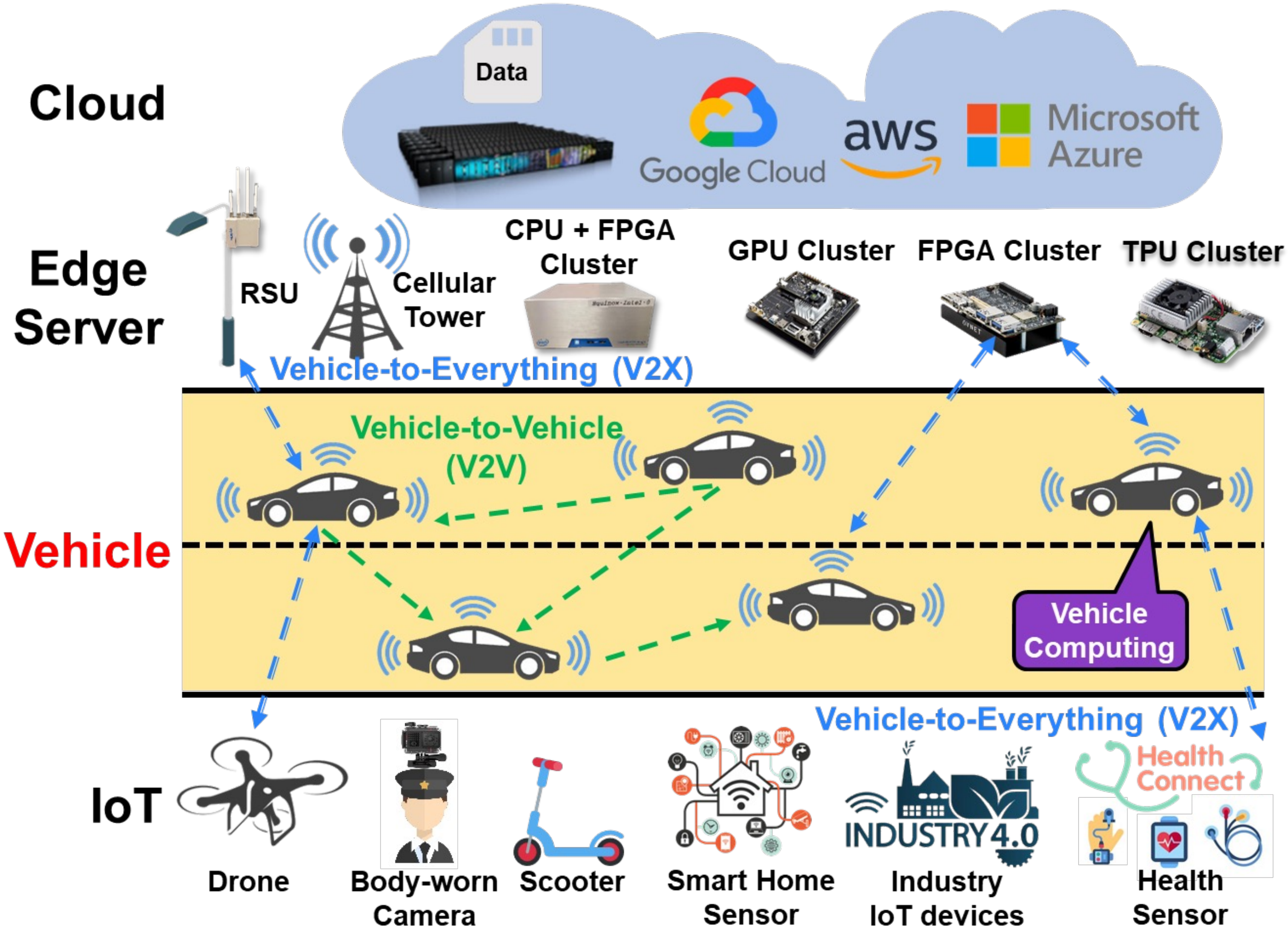


The Emergence of Vehicle Computing

The 4-Tier **Vehicle Computing** Paradigm

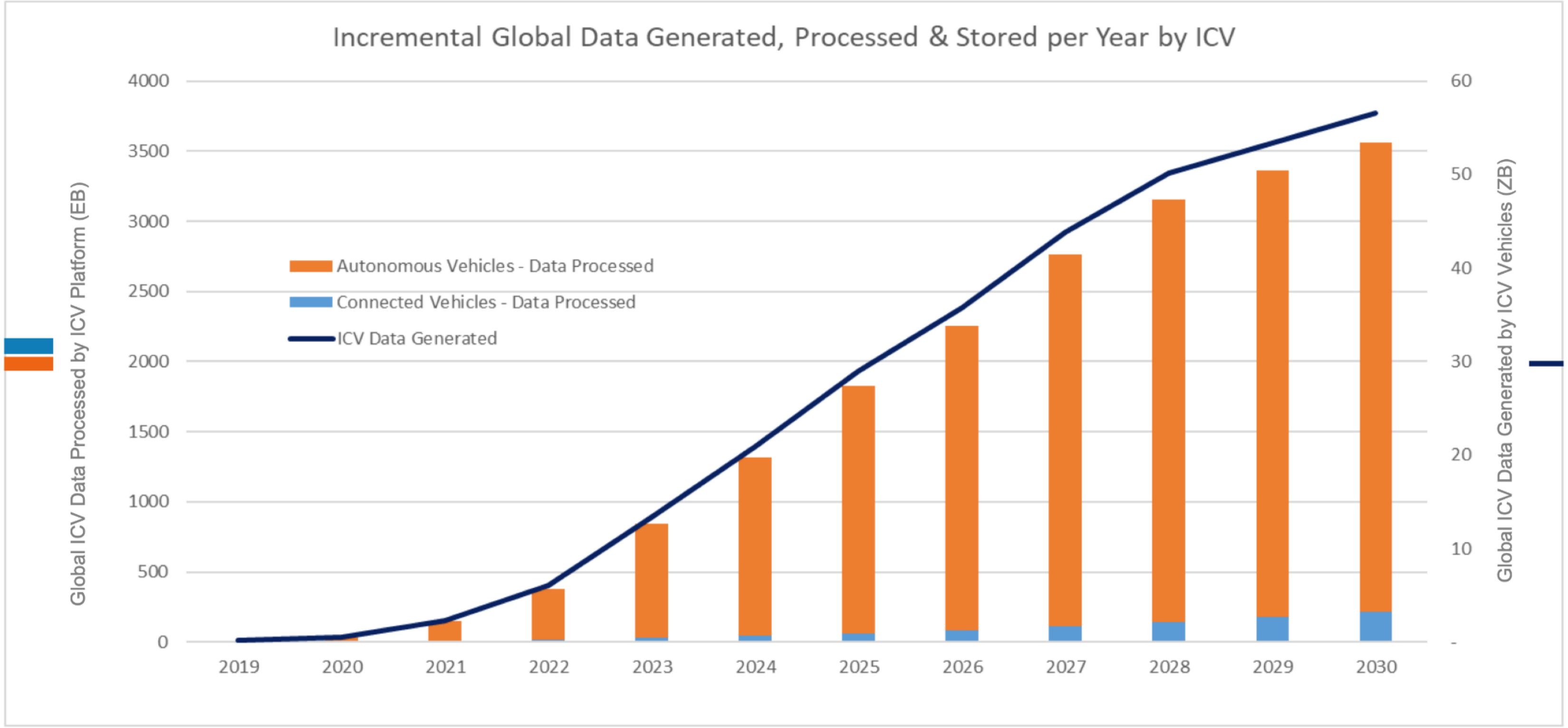
Computing on CVs based on data from:

- In-vehicle sensors,
- Surrounding **connected devices**

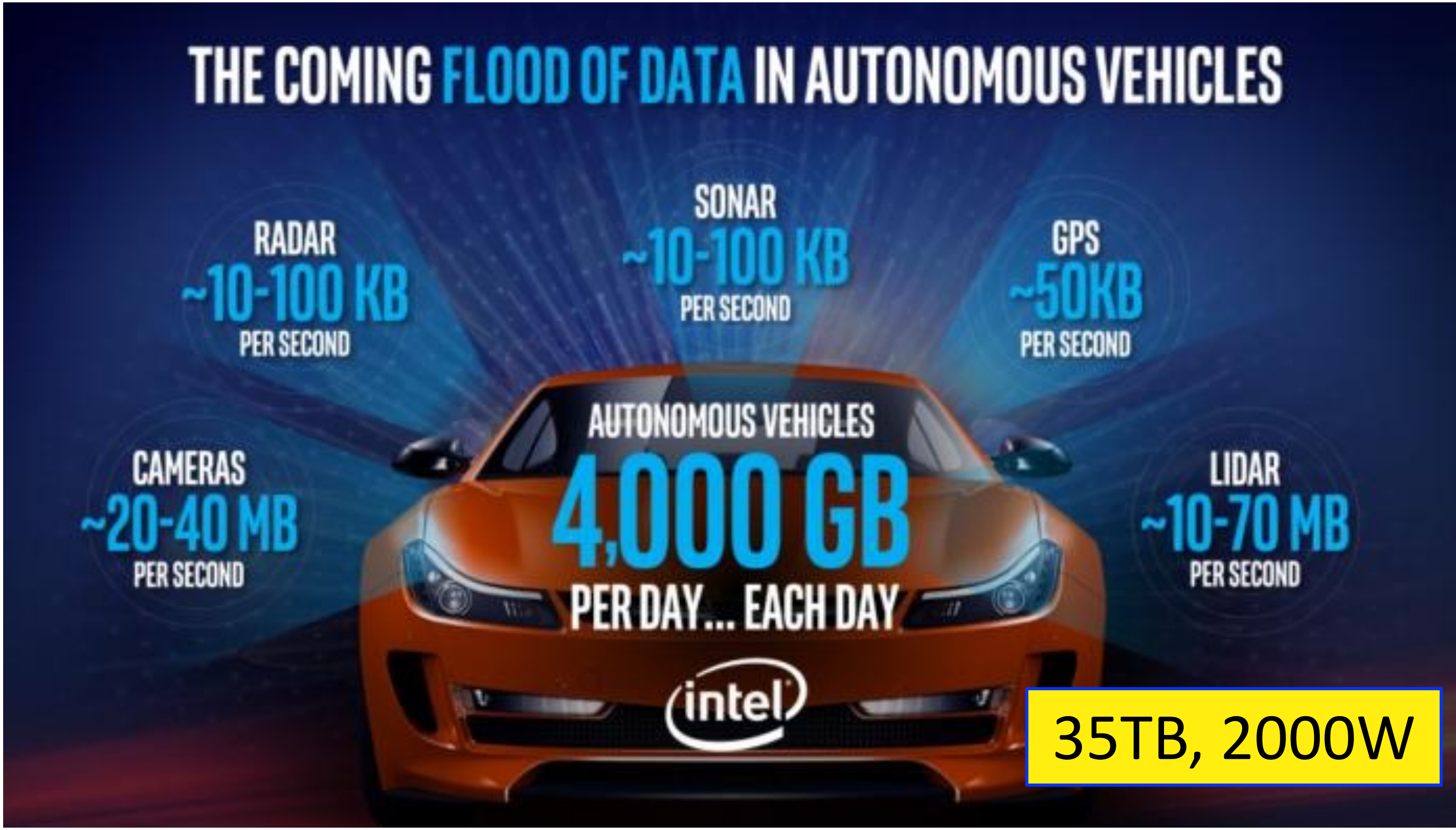


Data Generated by CAVs

ICV Represents Over 17% of Global Data Generated by 2025



Autonomous Vehicles

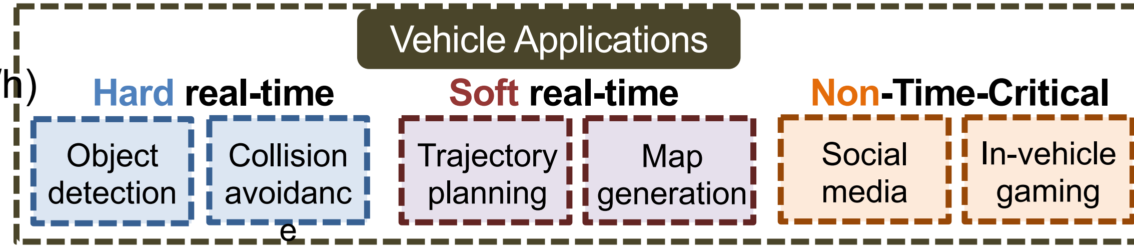


35TB, 2000W

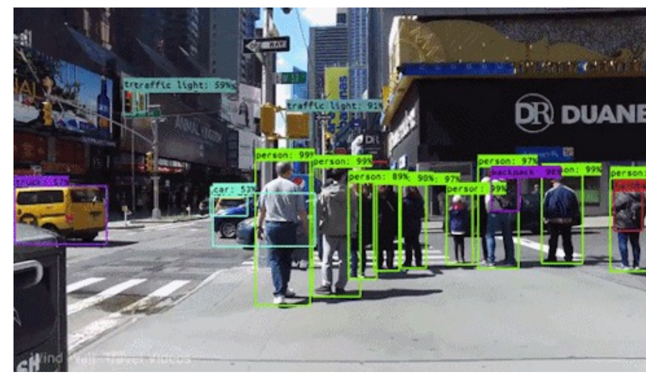
Credit: Intel

Challenge #1: Computation Latency

- **Time-sensitive** services
  - Response Time **< 90 ms** (40 km/h)
  - Computing Latency **<164ms** (avoid an obstacle at 5m away)



- **Vehicle data & model size**
  - Single CAV in urban: **40 TB** data / eight hours of driving
  - CAV fleets on highway: **280 PB** data
  - Increased model complexity



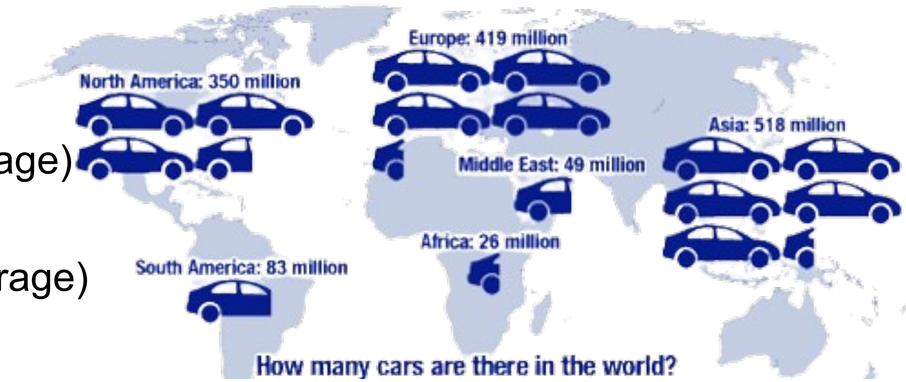
Reference: https://towardsdatascience.com/how-do-self-driving-cars-see-13054aee2503

- **Computation-constrained vehicles**
  - Traditional non-luxury vehicle: **$30K**
  - CAV: **$250K**
  - Sensors and computing platform: **two-thirds** of the total price

Goal: accelerate the inference speed of time-sensitive vehicle applications

Challenge #2: Transmission Costs

- **Transmission**
  - **Uplink:** data
    - 8GB data per vehicle, per day (on average)
  - **Downlink**: software/firmware update
    - 500MB per vehicle, per update (on average)
    - Update frequency: once per quarter



1.4 billion vehicles globally in 2022
(21% vehicles are in U.S.)

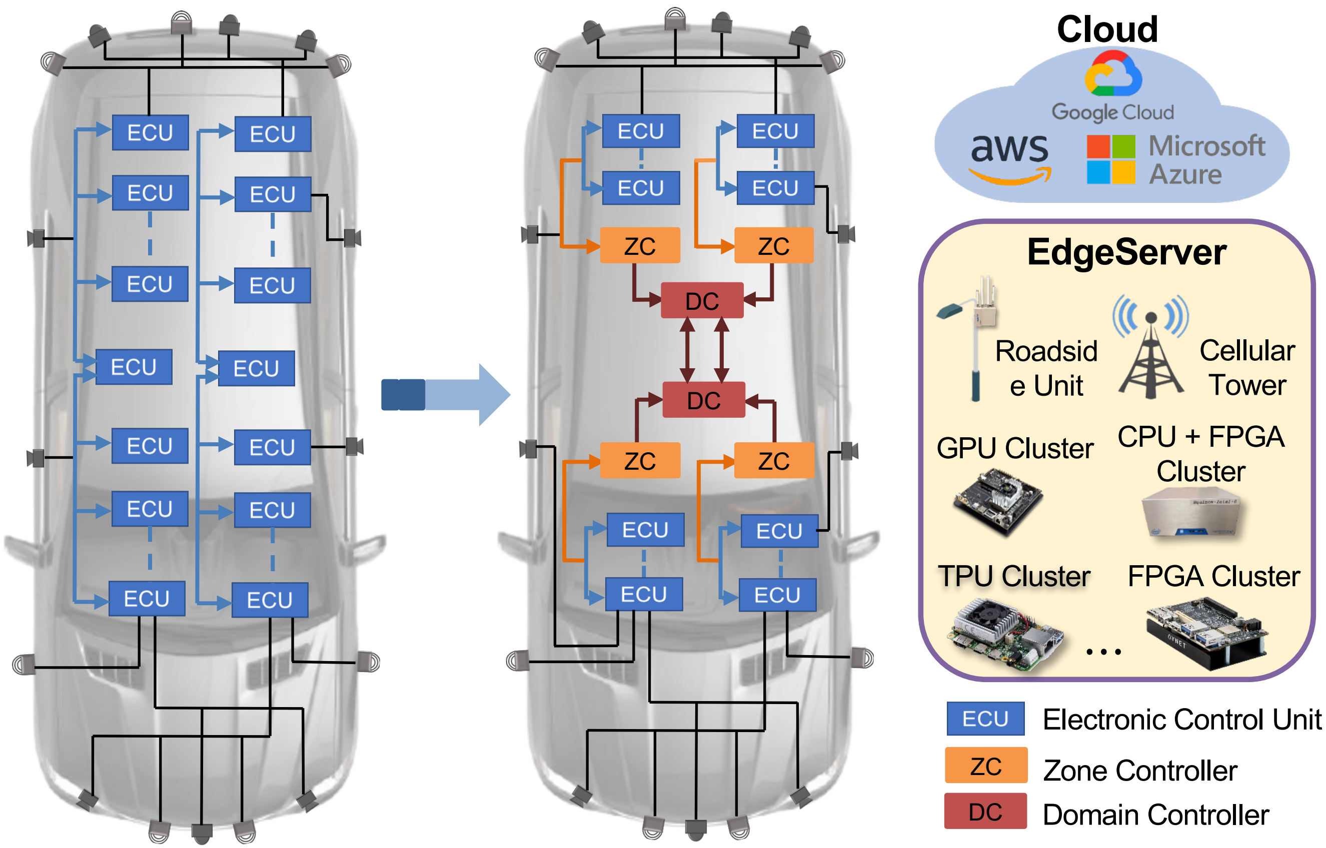
- **Transmission costs**
  - **Cost per usage**: 1 GB of mobile data worldwide: **$8.53** (**$12.37 in U.S**.)
  - **Unlimited prepaid data plan**: **$20** per month (AT&T, Chevy)

The cost of data transmission for a **10-million vehicle** fleet can reach over **20 PB** of data and cost over **$1 billion**, every year!

Enterprises can expect a **10 to 30% reduction** in costs from using Edge Computing.

Credit: https://hedgescompany.com/blog/2021/06/how-many-cars-are-there-in-the-world/

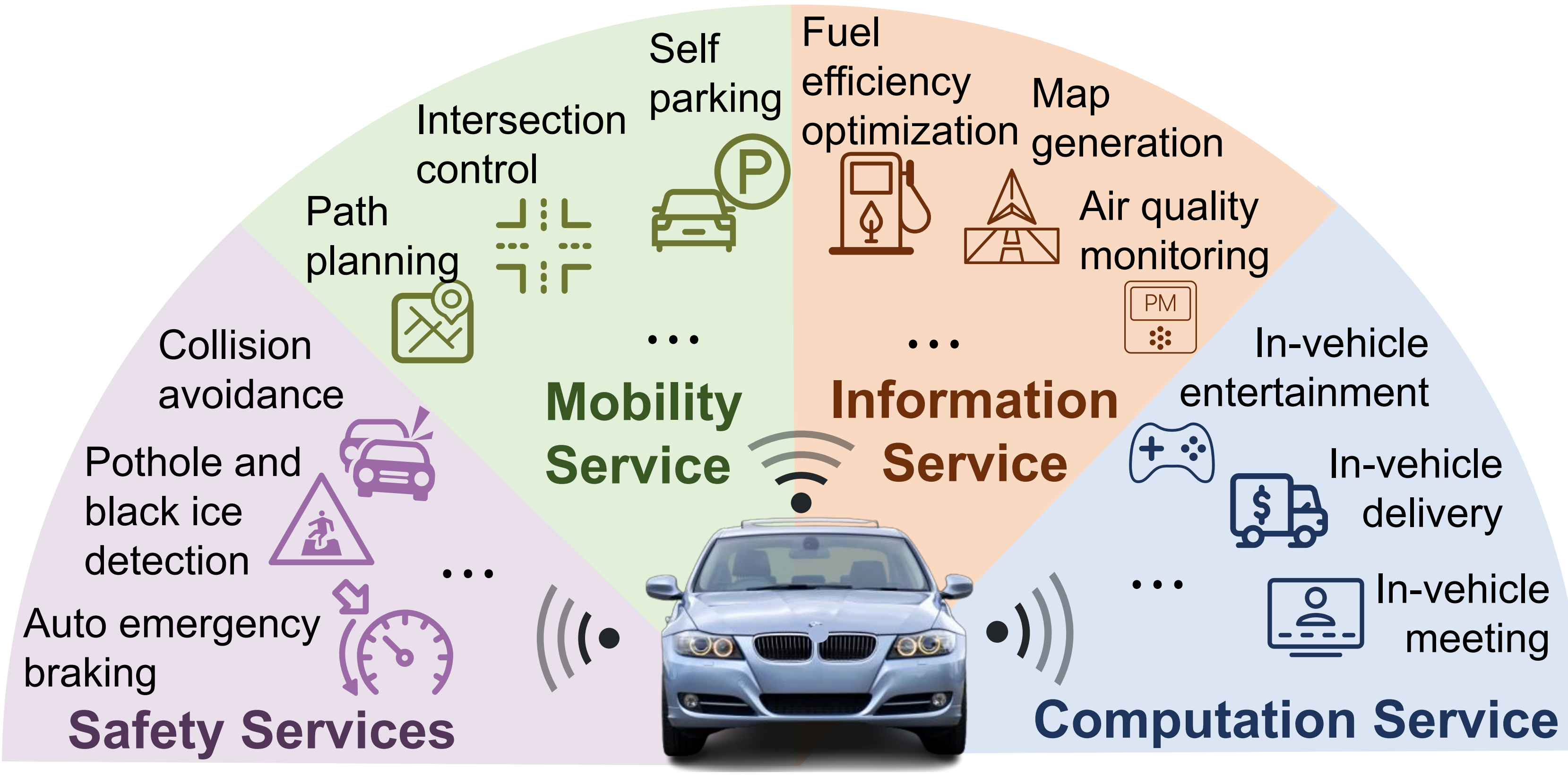
The Evolution of Automotive Computing System



Traditional Architecture **Software-Defined Architecture**

- **Shortcomings of traditional architecture:**
  - Difficult to deploy diverse computation-intensive applications.
- **Advantages of software-defined architecture:**
  - Simplifies vehicles' system interconnection
  - Makes the deployment of software to both ZCs and DCs possible

Software Defined Vehicles



Vehicles serve as both a sensor and a service producer and consumer.

Challenges in Vehicle Computing

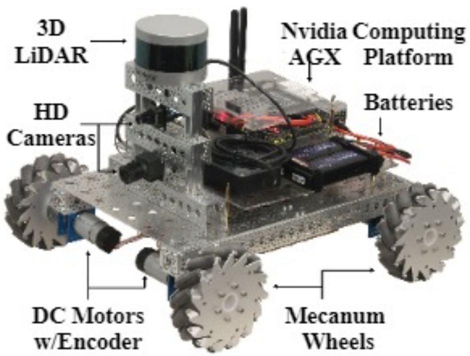
- Benchmarking and workload
- V2X communication
  - E.g., C-V2X, 5G/6G, WIFI
- Programmability (decomposition)
  - E.g., Novel programming model
- Runtime support and scheduling
  - E.g., automatically partition and deployment
- Energy consumption
  - E.g., computing, communication, sensing
- Security and privacy
  - E.g., trusted edge servers, Privacy-preserving
- End-to-end optimization
  - E.g., Communication/Computation/Control/Cost
- Business model
  - Automotive/Physical Infrastructure/Telecom/Cloud?
  - Deployment/Incentives

Roadmap

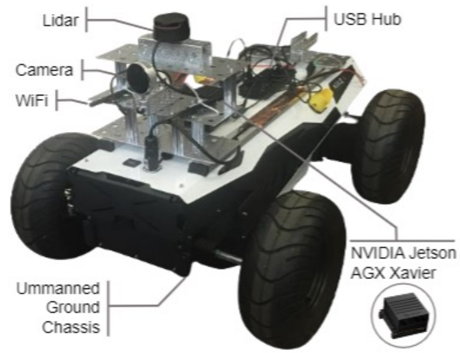
- Why Vehicle Computing?

☞ Research activities @CAR Lab

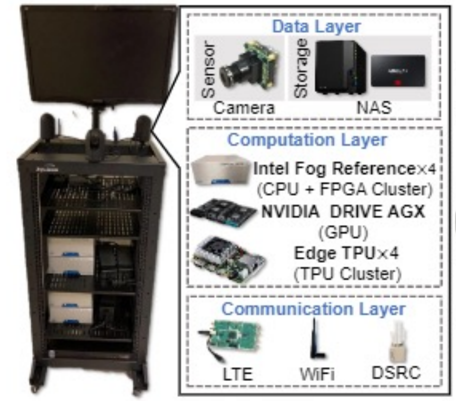
Research Platforms



HydraOne



Zebra



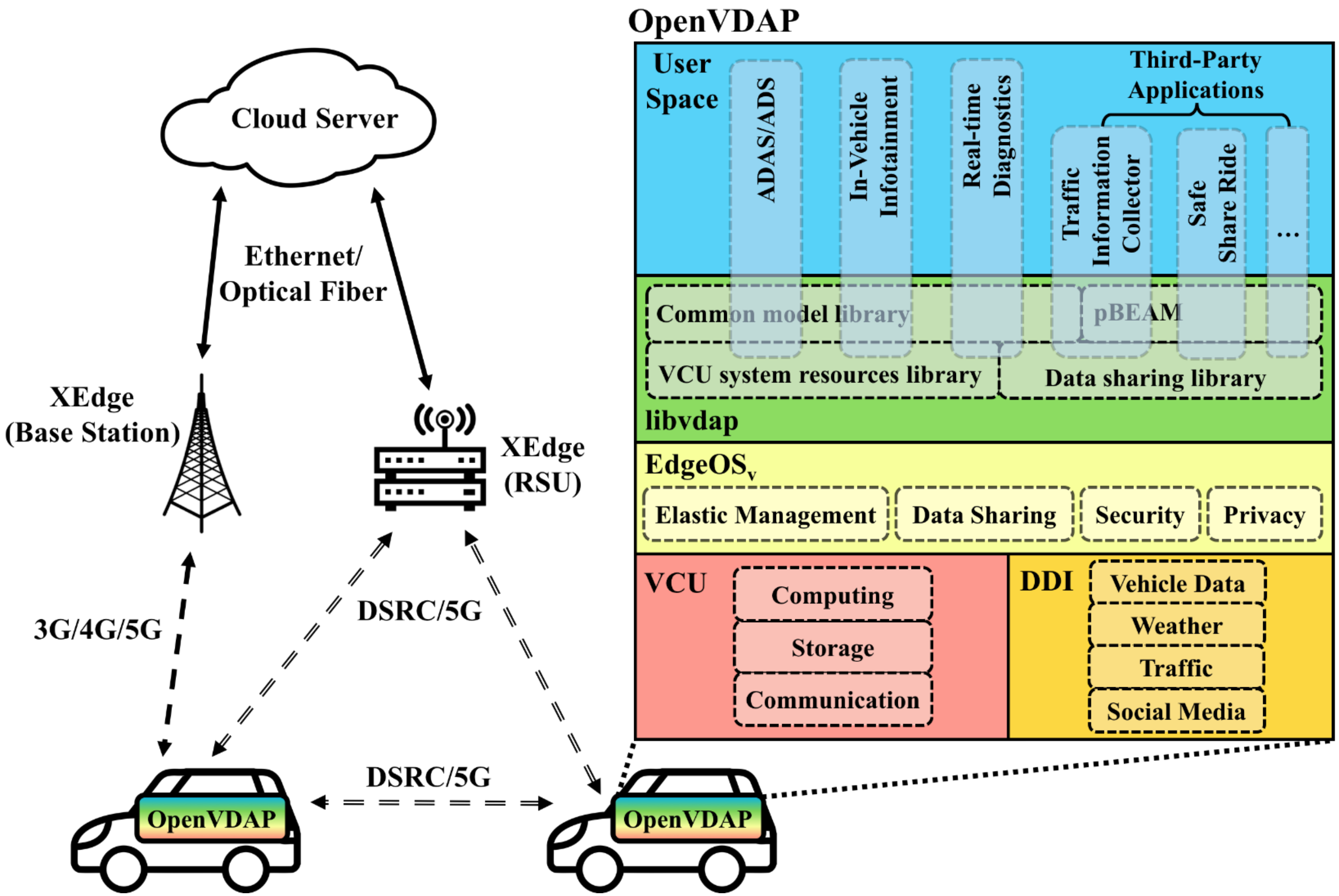
Equinox

ZebraT

Hydra

OpenVDAP for SDV (ICDCS'18)

- **Open V**ehicular **D**ata **A**nalytics **P**latform

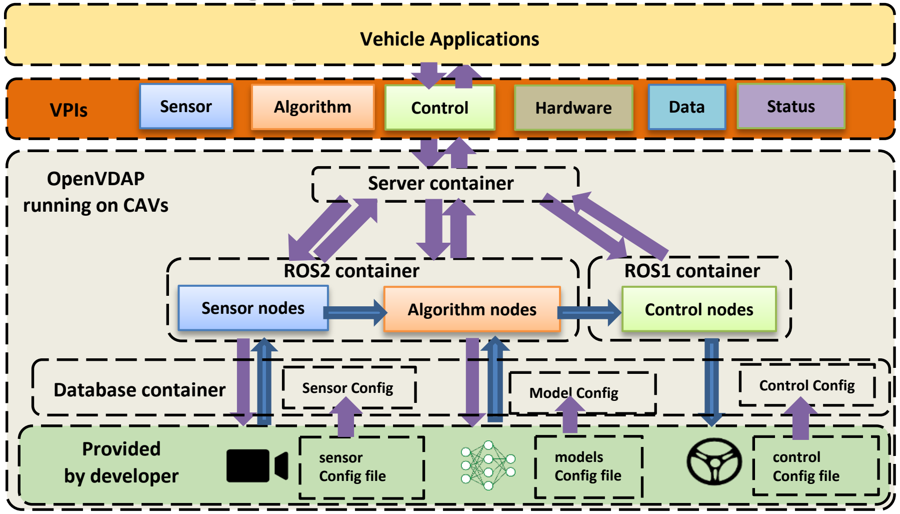


Vehicle Programming Interfaces

- Why VPIs？
  – No need the knowledge of vehicles, sensors, and communications
  – Only focus on application logic
  – Programming with *less* code
- Key VPIs design

VPI Types	VPI examples	Operations
Data	*vpi.data.getCameraData (front)*	get front camera data
	vpi.data.getSpatData()	get SPaT data from infrastructure
Control	*vpi.control.setTwist(msg)*	Set Twist command to CANbus
	vpi.control.setWiper(front, params)	Set wiper with params
Algorithm	*vpi.algorithm(*camera_front, e2e_lane_keeping_model*)*	Run end-to-end lane keeping model using front camera data
	vpi.algorithm([[camera_front_left,camera_front_right],lidar_top], [e2e_lane_keeping_model, collision_avoidance_model], test_case)	Run multi algorithms on test case
...	...	...

- The big picture

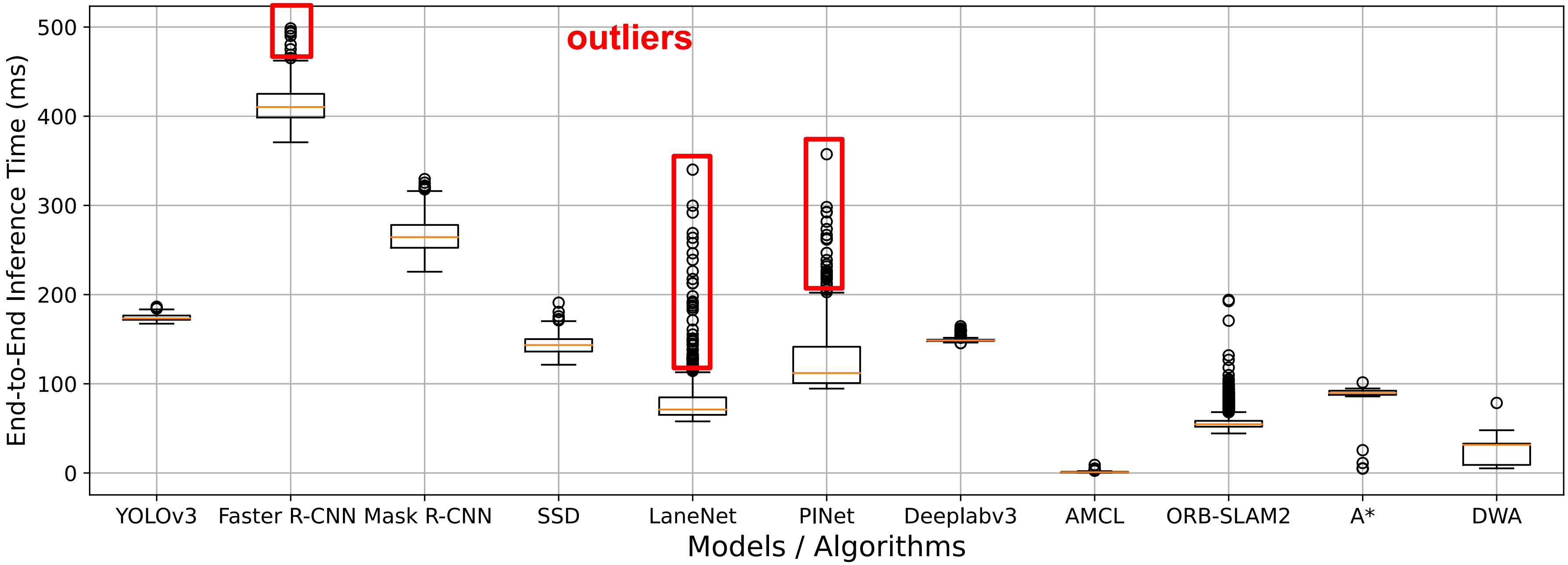


- An example

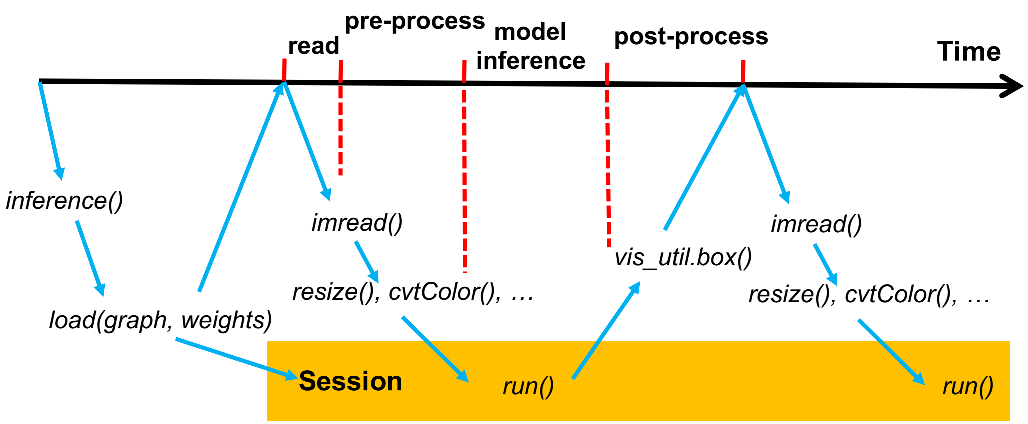
Lane keeping demo with 3 lines of code

```
import vpi
control_msg = vpi.algorithm(camera_front, e2e_lane_keeping_model)
vpi.control.setTwist(control_msg)
```

DNN Inference Time Variations in AVs



Timeline Analysis:



Potential variabilities:

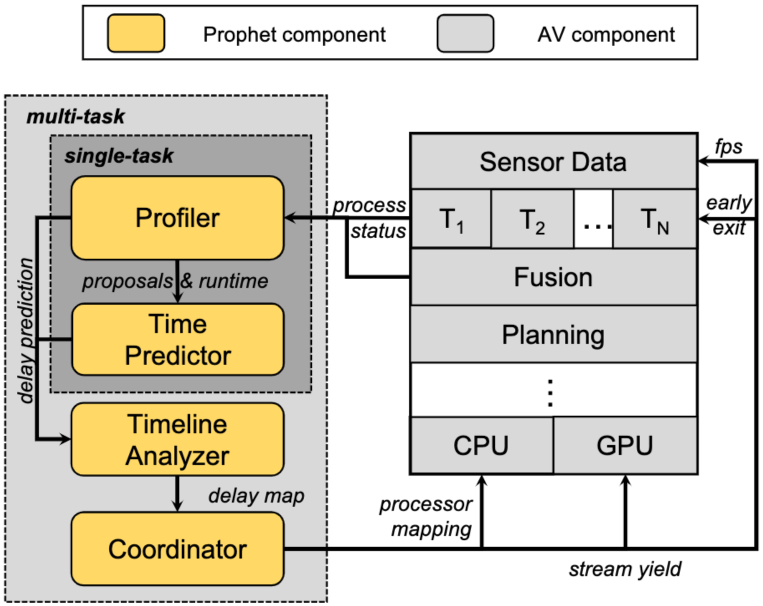
- **Read:** data, I/O methods
- **Pre-process:** data, hardware
- **Model inference:** model type, runtime, hardware
- **Post-process:** data, hardware

Six insights are derived in understanding the time variations for DNN inference.

Prophet: A Predictable Real-time Perception Pipeline for AVs (RTSS'22)

Two Insights from empirical study:

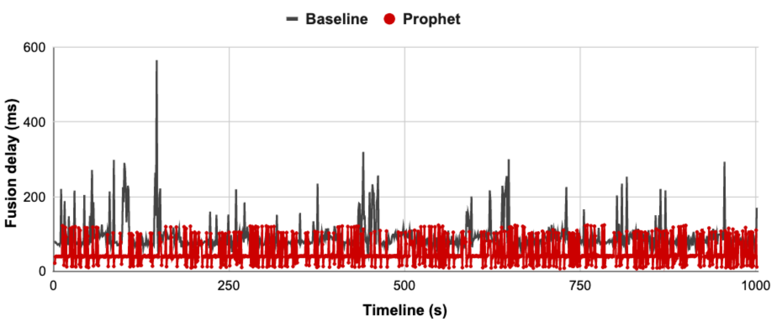
1. In silo mode, DNN's **structure** and the **runtime configurations** impacts the inference time variations.
2. In multi-tenant mode, proper **task coordination** is the key to addressing the time variations issue.



Inference time prediction:

Model	Real (ms)	Predicted (ms)	MAE (ms)	Accuracy (%)
Faster R-CNN	32.18	32.17	0.33	**98.99**
LaneNet	15.27	15.24	0.99	**94.03**
PINet	25.32	23.72	2.31	**91.68**

Perception system fusion delay:

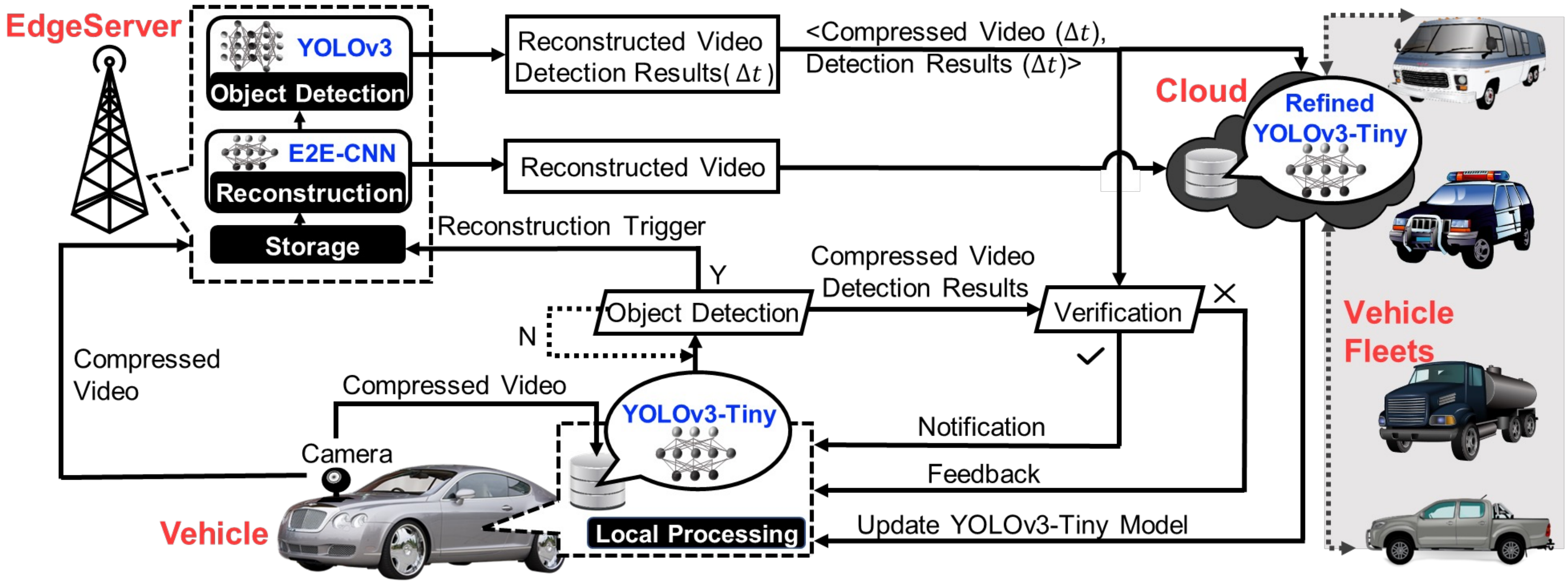


Key ideas:

- Predict inference time based on the **intermediate results** (proposals, raw points);
- **Early-exit** inference if the inference time is predicted to miss the deadline

Deadline miss rate:
5.4% (baseline) → 0.087% (Prophet)

Vehicle-Edge-Cloud Framework (SEC'20)



1) Vehicle

- *Energy-efficient network:* make timely computation on compressed data

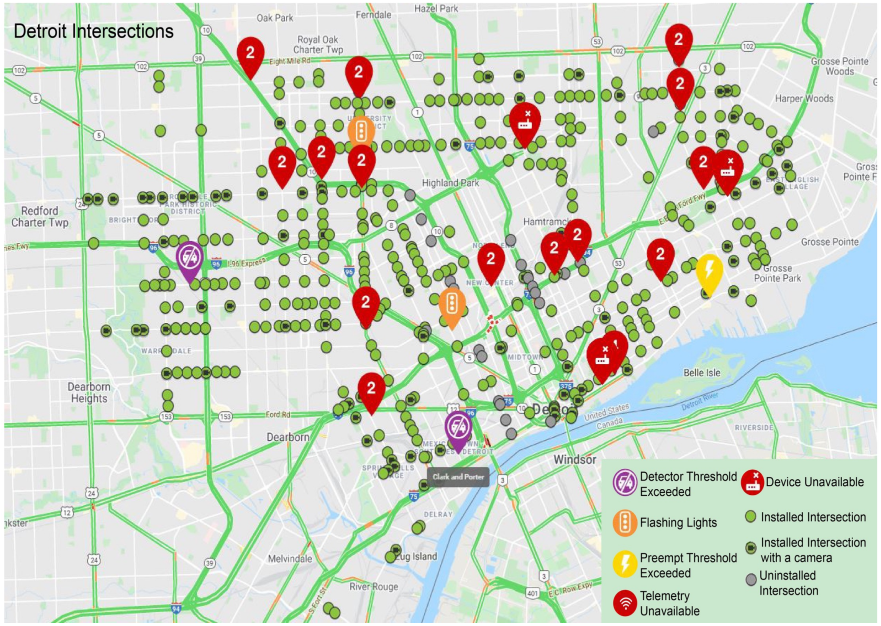
2) EdgeServer

- *Reconstruct* high-speed data with a triggered event
- *Verify* the detection results of the vehicle and send notifications

3) Cloud

- Aggregates all useful information
- Big data analysis: traffic control and path planning

Collaborators and Partners



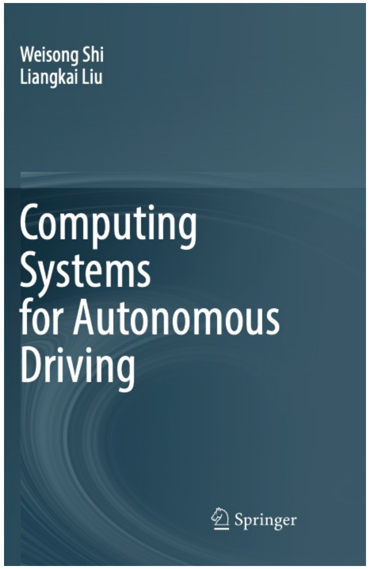
Summary

- *Vehicle computing era is coming*
- A lot of opportunities
  - Applications
    - CAV applications
  - Architecture/storage
  - Machine learning
  - Security/privacy
  - Systems/networking/communication
  - Tools
  - 4C Optimization

Additional Information

http://thecarlab.org

weisong@wayne.edu



Liangkai Liu, Sidi Lu, Ren Zhong, Baofu Wu, Yongtao Yao, Qingyang Zhang, Weisong Shi, *Computing Systems for Autonomous Driving: State-of-the-Art and Challenges*, **IEEE Internet of Things Journal, Vol. 8, No. 8, April 2021**.

Sidi Lu and Weisong Shi, *The Emergence of Vehicle Computing*, **IEEE Internet Computing Magazine**, May/June 2021.