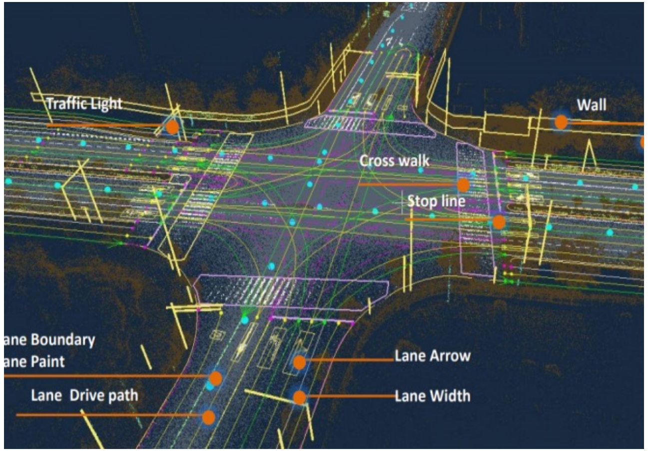
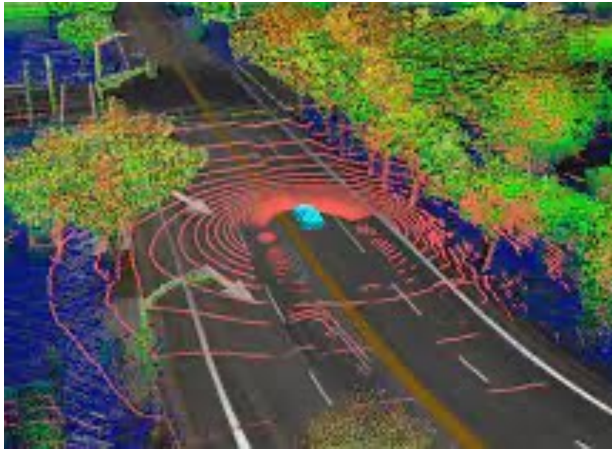
CoMap: Proactive Provision for Crowdsourcing Map in Automotive Edge Computing

Yongjie Xue, Yuru Zhang, Qiang Liu, *University of Nebraska-Lincoln*
Dawei Chen, **Kyungtae (KT) Han**, *Toyota InfoTech Labs*

High-Definition Map

- High-definition (HD) map enables Autonomous Driving and Advanced Driver Assistance Systems (ADAS)
  - Accurate and high-precision presentation of the roads
  - Autonomous Driving and ADAS rely on HD maps for localization, e.g., SLAM

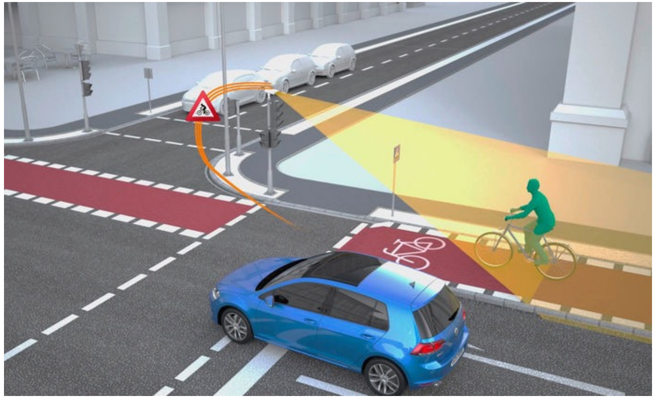




High-Definition Map

❖ HD map needs up-to-date information

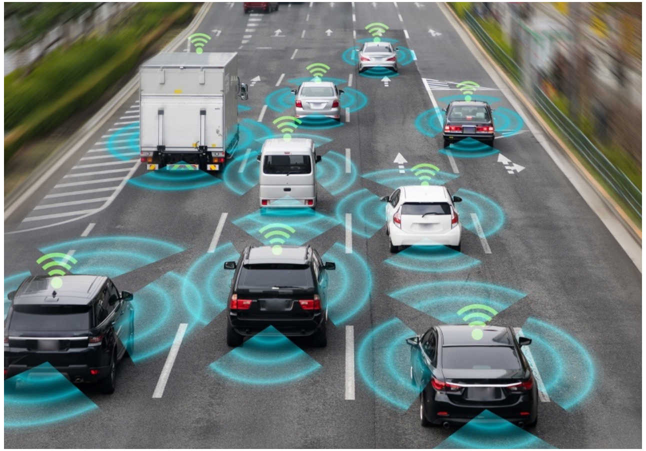
- Transient information on the road, e.g., constructions and accidents
- Infrastructural sensors have limited coverage and angles

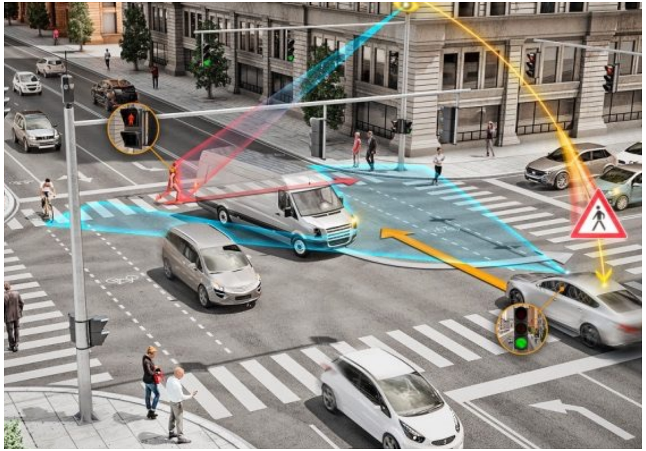


Connected and Automated Vehicles (CAV)

❖ CAVs connect vehicles wirelessly with Edge Computing

- Allow information sharing and vehicle collaborations
- Crowdsourcing data from rich sensors in CAVs for updating HD map





TOYOTA
InfoTech
Envisioning Mobility

Demands

- Enormous UL/DL radio transmission needs
  - Raw sensor data can be up to 100Mbps per CAV
  - High operating expenses (OPEX) for service providers
- Fast-changing network dynamics
  - High-velocity of vehicles, e.g., channel condition and traffic
  - Complicated resource demands

SLAM
>100Mbps

Speed
>30kph

* Ahmad, F., et. al., Carmap: Fast 3d feature map updates for automobiles. NSDI 2020

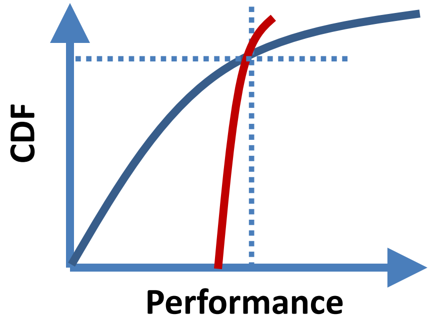
TOYOTA InfoTech
Envisioning Mobility

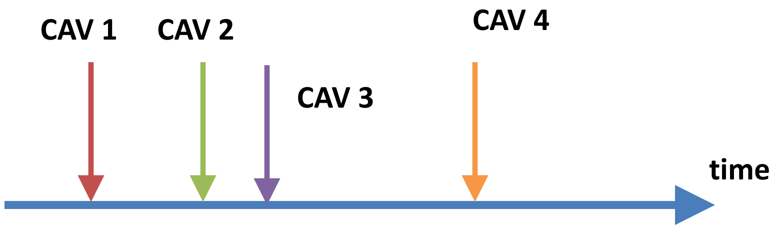
Existing Work

❖ Focus on the average performance
- Long-tail latency provides very limited information in HD map

❖ Centralized control plane
- Delayed optimized variables worsen the latency performance

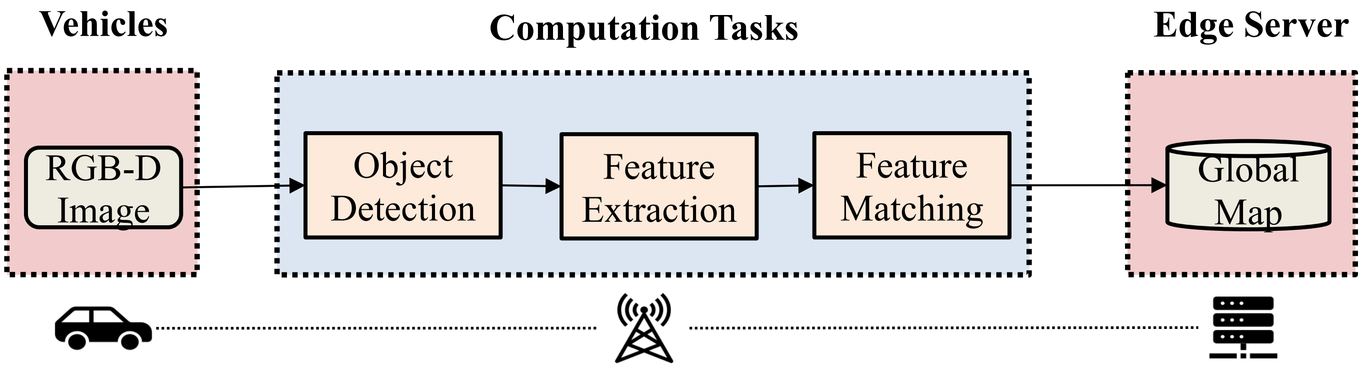
❖ Static resource allocation
- Cannot fit asynchronous offloading in real world





CoMap

- **Objective**
  - Optimize **network** and **compute** resource usage
- **Insight 1: Adaptive vehicular offloading**
  - The more pre-processes onboard, the fewer data to be transmitted
- **Insight 2: Learning resource reservation/demand**
  - Satisfy the latency requirement of vehicular offloading



System Model

❖ System Consideration

- Multiple base stations and edge servers, and CAVs
- Flexible partition of computation tasks between [0, 1] (a_n)

❖ End-to-end latency

- Local processing delay: Computation over vehicle capacity
- Uplink delay: Data size over wireless data rate
- Edge processing delay: Computation over edge capacity
- Static delay: Overhead in protocol and others

$$L_n = \boxed{f(a_n)/F_n} + \boxed{g(a_n)/(x_n \cdot E_n)} + \boxed{h(a_n)/y_n} + \boxed{D_n}$$

TOYOTA InfoTech
Envisioning Mobility

The Resource Provision Problem

- **Objective function:** Optimize the monetary cost for all CAVs
- **Optimization variables**
  - Computation partition, radio bandwidth, edge computation capacity
- **Constraint:** Percentile latency requirements

$$\max_{\mathcal{A},\mathcal{X},\mathcal{Y}} \quad \sum_{t=0}^{T} \sum_{n=0}^{N} \left(\boxed{x_n^t / B} + \eta \cdot \boxed{y_n^t / G} \right) \tag{2}$$

$$s.t. \quad \boxed{Pr\left(\mathcal{L} \le H\right) \ge p,} \tag{3}$$

$$0 \le \sum_{n \in \mathcal{N}} x_n^t \le B, \forall t \tag{4}$$

$$0 \le \sum_{n \in \mathcal{N}} y_n^t \le G, \forall t \tag{5}$$

$$0 \le a_n \le 1, \forall n, t \tag{6}$$

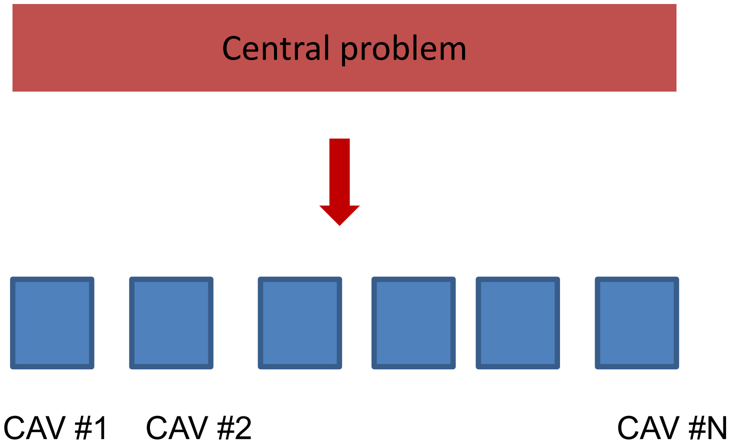
Technical Challenges

- Needs to be fully decentralized
  - No information on other CAVs
- Probabilistic resource demands
  - Transmission data size and computation complexity are not fixed, but stochastic
- Temporal resource allocation
  - The resources allocation is changed over time

Our Solution (1/4)

❖ Reduce the problem

- First, optimize for the ego CAV
- We will handle the inter-CAV constraints, later



$$
\begin{aligned}
\min_{\mathcal{A}_n,\mathcal{X}_n,\mathcal{Y}_n} \quad & \sum_{t=0}^{T} \left(\boxed{x_n^t/B} + \eta \, \boxed{y_n^t/G}\right) \\
s.t. \quad & \boxed{Pr\left(L_n \leq H\right) \geq p,} \\
& 0 \leq x_n^t \leq B, \forall t \\
& 0 \leq y_n^t \leq G, \forall t \\
& 0 \leq a_n \leq 1, \forall n, t
\end{aligned}
$$

Our Solution (2/4)

❖ Predict resource demand under the extra context

- We find they are related to some environmental context, e.g., CAV id, CAV location, sensor rotation
- We use Bayesian neural network for context-based prediction

Gaussian process → Scalability is poor, when dataset is large, e.g., 1000+

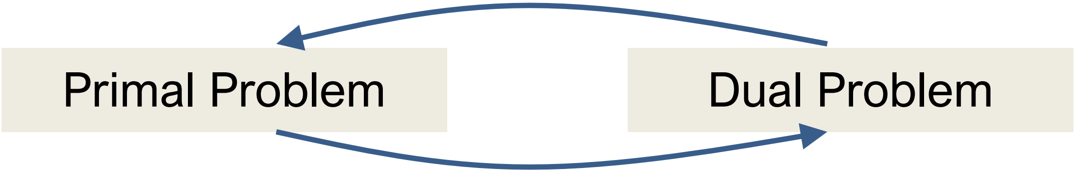
DNN → Scalable, but deterministic output, fails in percentile latency

Bayesian Neural Network → We will BNN to predict.

TOYOTA InfoTech
Envisioning Mobility

Our Solution (3/4)

- Solve the problem with the Lagrangian methods
  - We convert the stochastic problem into a deterministic one
  - First, an exhaustive search of the computation partition (discrete in practice)
  - Second, use the Lagrangian primal-dual method to derive the optimal solution



Our Solution (4/4)

- Infrastructural temporal balancing
  - Each CAV makes its resource request, which may over-request
  - We find that the task can always be completed as long as the total allocated resources are enough
  - Use **water-filling** like algorithm to balance the temporal resource usage

Time slot	1	2	3	4	5	6
user 1	2	1	1	1	1	1
user 2	0	6	0	0	0	0
user 3	0	0	0	0	0	6

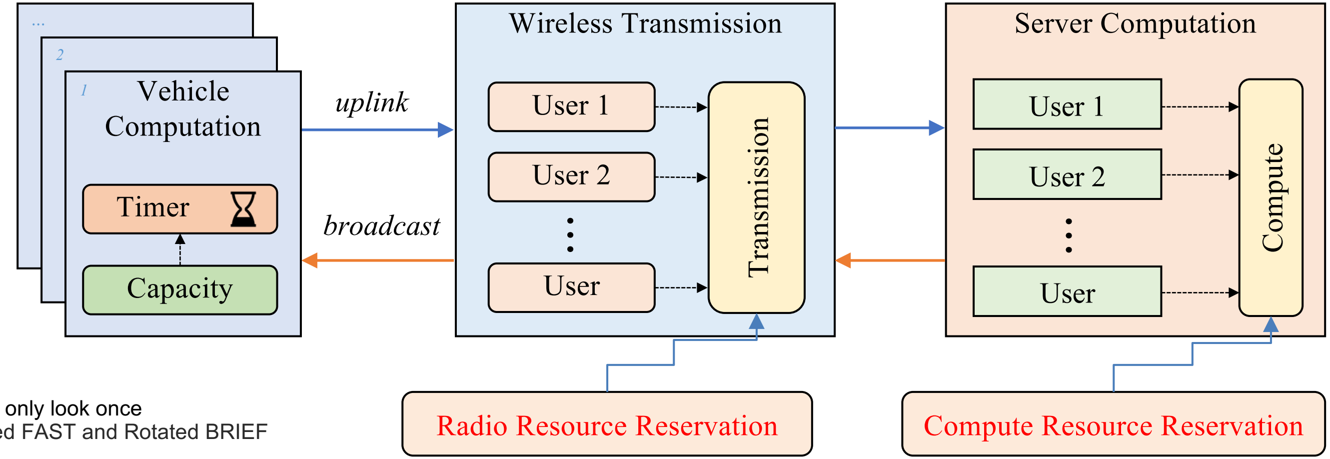
Simulation

❖ End-to-End Network Simulator

- Time-driven, with 5G UL/DL and queue-based edge computation
- Packet sizes and computing time are collected from real experiments/profiling

❖ Other Parameters

- YOLOv5 object detection, ORB feature extraction, brutal-force feature matching
- V2X-Sim Dataset, including 50 CAVs over 100 frames
- Radio bandwidth 10MHz, latency requirement is 100ms with 90^{th} percentile



* YOLO: You only look once
ORB: Oriented FAST and Rotated BRIEF

Results (1/3)

❖ Simulation Results – Overall Performance

- Our solution achieves the given percentile requirements, i.e., Prob(L<100ms)>90%
- Our solution obtains the lowest resource usage with assured requirements

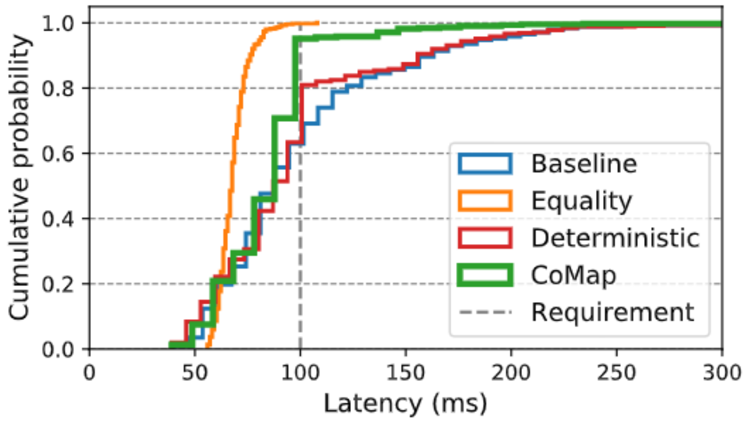


Fig. 1: Latency performance of algorithms

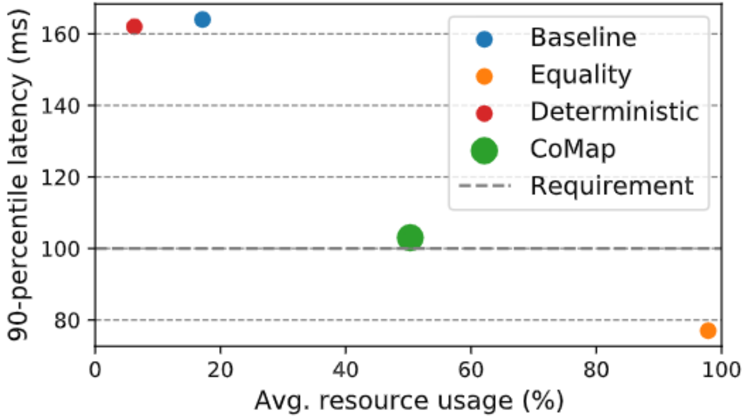


Fig. 4: Performance of algorithms

Results (2/3)

❖ Simulation Results – Temporal Resources

- Our solution achieves on-demand resource allocation in time domain
- More in the radio, and relatively less in computation.

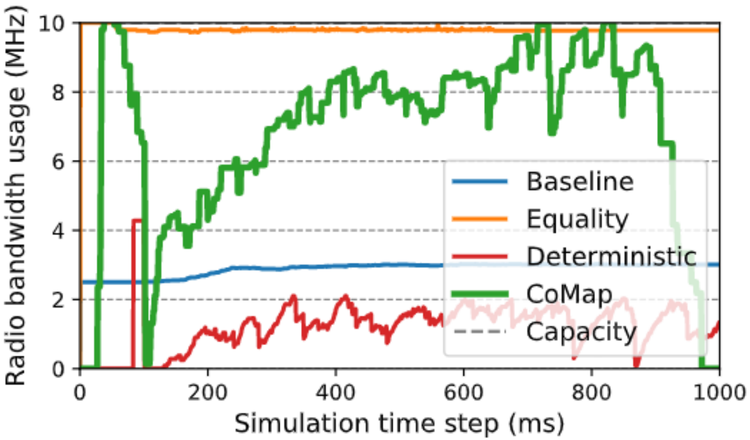


Fig. 2: Radio resource allocation

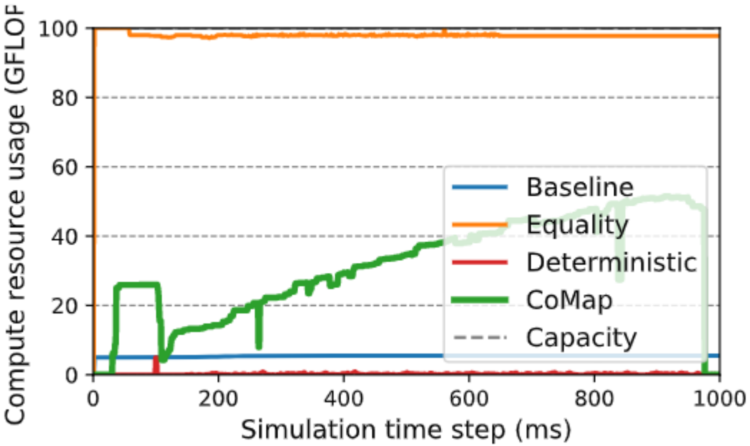


Fig. 3: Compute resource allocation

Results (3/3)

❖ Simulation Results – Scalability

- Our solution maintains the percentile requirements at scale
- Large reduction on resource usage as compared to Equality

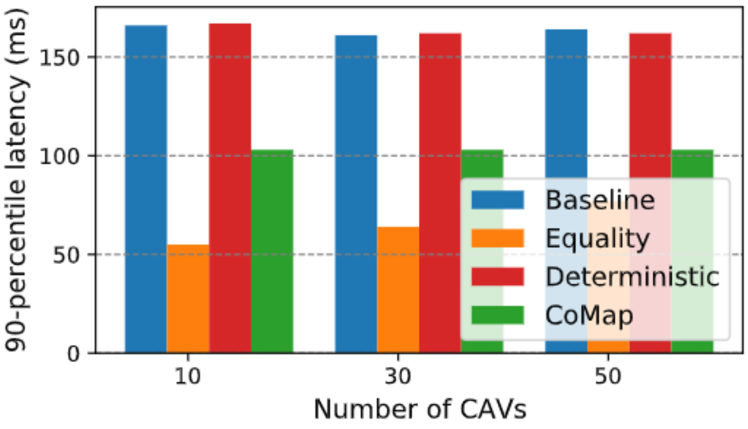


Fig. 5: Latency under different traffic

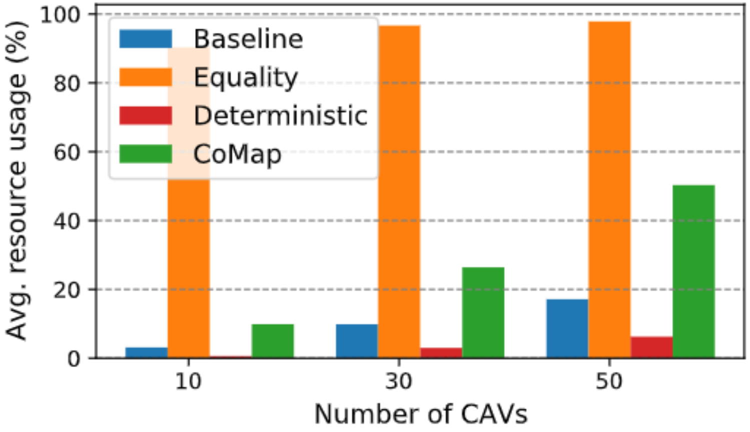


Fig. 6: Usage under different traffic

Summary

- We proposed **a new provisioning method** for new crowdsourcing HD maps in automotive edge computing
- To achieve deterministic latency performance, we design a new algorithm that **proactively** allocates temporal resource
- We developed an **end-to-end network simulator** with traces from experimental profiling
- The proposed algorithm shows the **overall performance is better** than baseline algorithms.

TOYOTA InfoTech
Envisioning Mobility

Thank You!

kt.han@toyota.com

TOYOTA InfoTech
Envisioning Mobility

TOYOTA InfoTech
Envisioning Mobility