

# AutoCast: Scalable Infrastructure-less Cooperative Perception for Distributed Collaborative Driving

Hang Qiu  
University of Southern California  
Los Angeles, California, USA  
hangqiu@usc.edu

Po-Han Huang  
University of Southern California  
Los Angeles, California, USA  
pohanh@usc.edu

Namo Asavisanu  
University of Southern California  
Los Angeles, California, USA  
namo@usc.edu

Xiaochen Liu  
University of Southern California  
Los Angeles, California, USA  
liu851@usc.edu

Konstantinos Psounis  
University of Southern California  
Los Angeles, California, USA  
kpsounis@usc.edu

Ramesh Govindan  
University of Southern California  
Los Angeles, California, USA  
ramesh@usc.edu

## Abstract

Autonomous vehicles use 3D sensors for perception. Cooperative perception enables vehicles to share sensor readings with each other to improve safety. Prior work in cooperative perception scales poorly even with infrastructure support. AUTOCAST<sup>1</sup> enables scalable infrastructure-less cooperative perception using direct vehicle-to-vehicle communication. It carefully determines which objects to share based on positional relationships between traffic participants, and the time evolution of their trajectories. It coordinates vehicles and optimally schedules transmissions in a distributed fashion. Extensive evaluation results under different scenarios show that, unlike competing approaches, AUTOCAST can avoid crashes and near-misses which occur frequently without cooperative perception, its performance scales gracefully in dense traffic scenarios providing 2-4x visibility into safety critical objects compared to existing cooperative perception schemes, its transmission schedules can be completed on the real radio testbed, and its scheduling algorithm is near-optimal with negligible computation overhead.

## CCS Concepts

• Computing methodologies → Cooperation and coordination.

## Keywords

Cooperative Perception, Autonomous Cars, V2V Communication

## ACM Reference Format:

Hang Qiu, Po-Han Huang, Namu Asavisanu, Xiaochen Liu, Konstantinos Psounis, and Ramesh Govindan. 2022. AutoCast: Scalable Infrastructure-less Cooperative Perception for Distributed Collaborative Driving. In *The 20th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '22)*, June 25-July 1, 2022, Portland, OR, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3498361.3538925>

<sup>1</sup> AUTOCAST is available at <https://hangqiu.github.io/AutoCast/>

This material is based upon work supported by the National Science Foundation under Grant No. 1956445. This work is in part supported by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

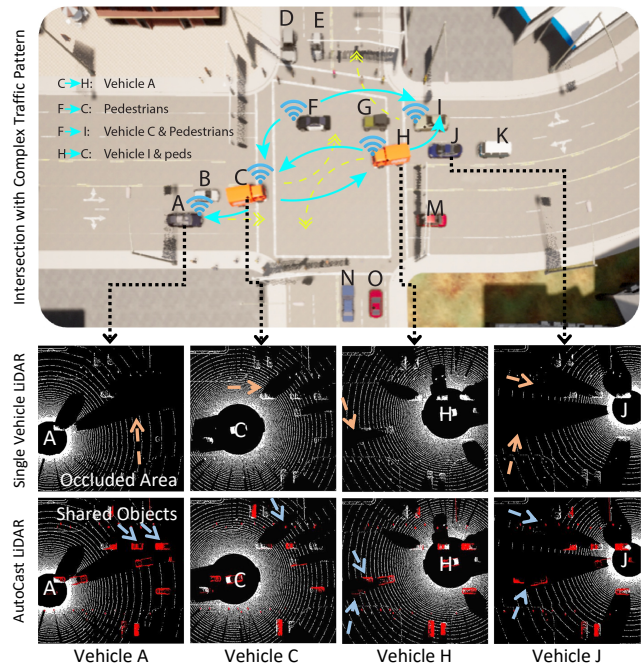
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobiSys '22, June 25-July 1, 2022, Portland, OR, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9185-6/22/06.

<https://doi.org/10.1145/3498361.3538925>



**FIGURE 1:** AUTOCAST enables multi-vehicle cooperative perception in a busy intersection. The top graph shows that AUTOCAST orchestrates vehicles to selectively share useful information (light blue arrows) about occluded objects, which the receiver vehicles cannot see, but may affect the receivers trajectories (yellow dashed arrow). The bottom graph shows the LiDAR perspectives of vehicle A, C, H, J: the upper row shows the invisible area (orange arrows) before sharing; the lower row shows the previous invisible objects (red points, blue arrows) visible after sharing.

## 1 Introduction

Autonomous driving technology has made great strides in transforming transportation. To be socially acceptable and widely deployed, it needs to ensure reliability over a broad set of unusual traffic situations and corner cases [49], reaching or surpassing human driving safety levels (100M miles between fatalities [61]).

The reliability of today's autonomous driving systems is critically dependent on the accuracy of its perception component. However, 3-D sensors like LiDAR and stereo cameras suffer from line of sight limitations: other vehicles and traffic participants (pedestrians, bicyclists) can block a vehicle's view. For example, in a popular vehicular dataset (nuScenes [26]), 53.14% of over 204K labeled pedestrians and 59.43% of 483K annotated vehicles are occluded. The blind spots caused by occlusion can compromise the reliability

of object detection and downstream path planning in many driving situations, including left turns, lane changes, and overtaking.

**Cooperative perception.** To address this limitation, recent work [27, 41, 64, 78, 82] has proposed a novel new direction, *cooperative perception*. In this approach, vehicles cooperatively exchange sensor readings from 3D sensors to extend their visual horizon (Figure 1). The benefits of cooperative perception for autonomous driving systems are clear: a vehicle can make decisions much earlier than it otherwise might have been able to. Cooperative perception is a natural next step beyond earlier prior work that considered safety enhancements by having vehicles actively broadcast their location continuously over short range DSRC radios [47]; these approaches cannot capture passive participants (pedestrians and bicyclists), but cooperative perception can.

**Connected vehicles: enabling cooperative perception.** Prior work leverages recent advances in vehicular communication to enable cooperative perception. AVR [65] enables vehicles to *directly* exchange stereo camera point clouds via *vehicle-to-vehicle* (V2V) communication. EMP [82] exploits infrastructure support [22] to share non-overlapping segments of LiDAR point clouds via *vehicle-to-infrastructure* (V2I) communication, using an edge server as a *relay*. Using static infrastructure, V2I can achieve higher bandwidth compared to direct mobile V2V communication. For example, V2I using LTE [23] can achieve nominal bandwidths of up to 300 Mbps [2–4], whereas current commercial V2V products using DSRC<sup>2</sup> [13] can only achieve around 6 Mbps. LTE and Wi-Fi, which many modern vehicles have, increasingly support *direct* modes and wider channel bandwidths. However, WiFi-direct via 802.11n/ac or 60 GHz (ad) cannot adapt to the highly variable wireless channel between fast moving vehicles. Existing LTE-direct [39] chips for vehicular applications, tend to achieve around 10 Mbps<sup>3</sup>.

**Sharing point clouds is desirable but challenging.** Like EMP and AVR, we advocate for sharing raw points for cooperative perception, as opposed to processed information, such as bounding boxes or visual features. Sharing bounding boxes limits cooperative perception by the accuracy of the object detector deployed on the transmitter vehicle. Among popular object detectors (VoxelNet [83], PointPillars [51], CenterPoint [79]), the mean average precision (mAP) can vary by 8%. The other alternative is to use a neural network to extract and share features. This approach would *constrain innovation*: the receiver might wish to use the shared feature for many different purposes (object detection, drivable space segmentation, trajectory planning), and the transmitted features might limit the efficacy of these tasks. Even for object detection, prior work has shown that early fusion of shared point clouds can result in *higher accuracy* compared to late fusion of processed features [78].

Nevertheless, sharing raw points is challenging. While AVR and EMP have demonstrated feasibility, their scale is limited by the network bottleneck: AVR is limited to two vehicles using V2V communication; EMP scales up to six vehicles via V2I, but needs infrastructure support.

**Our focus: Scalable Infrastructure-less Cooperative Perception.** In this paper, we explore the problem of *scalable infrastructure-less cooperative perception*, which permits vehicles to share raw

sensor data in dense traffic scenarios without the dependency on edge servers. To motivate the problem, consider a busy intersection (Figure 1) with complex traffic dynamics where no infrastructure support is available: people and bicyclists crossing the street, traffic waiting to make a turn, together with traffic flowing in the direction of the green light. In such a scenario, there may be tens to hundreds of traffic participants; if each vehicle could still share information about participants that it sees, other vehicles would have more complete information to plan better trajectories. Without infrastructure support, our approach faces two challenges: fitting the shared data into the narrower V2V bandwidth, and coordinating and scheduling transmissions to avoid packet collision.

**Sharing point clouds via V2V channel is feasible.** The closest prior work, EMP [82], demonstrated that it is possible to transfer, through V2I channel, up to six non-overlapping segments of point clouds (each of size 30–38 KB). For the purpose of autonomous driving and collision avoidance, instead of an entire segment, a vehicle can transmit point clouds of relevant objects in the scene. By emulating a 64-beam Velodyne LiDAR (which generates 2.2 M points/sec [12]) in photorealistic simulations of different driving scenarios (see §5), we have found that point clouds for a detected object (*e.g.*, a truck close by) have up to 200 points (38.4 kbits<sup>4</sup>, an order of magnitude smaller than EMP segments). Ideally, to enable a vehicle to track participants (especially fast-moving vehicles) precisely, each vehicle must receive (and make trajectory planning decisions on) point clouds at sub-second timescales. The finest *decision interval*, denoted by  $T_d$  is 100 ms, which is the interval at which the Velodyne LiDAR generates a frame [12]. At 10 Hz (as clocked in EMP, when  $T_d = 100\text{ms}$ ), a nominal V2V channel capacity of 10 Mbps can fit from 25 dense vehicle point clouds to hundreds of sparser (faraway) or smaller objects (pedestrians and cyclists). Compression can further increase that number.

**Distributed coordination and scheduling.** Enabling cooperative perception with infrastructure support can require a huge investment with limited coverage. For example, EMP [82] supports cooperative perception across only six vehicles every 100 m. Deploying such services in urban scenarios, either to regional road-side units (RSUs) or to aggregated edge clusters (*e.g.*, cellular towers), can be very expensive in terms of compute and cost. Moreover, using edge servers as relays unnecessarily duplicates transmissions, which wastes already scarce wireless bandwidth, and increases end-to-end (V2V) delivery latency. In areas not covered by such edge services, there is no scalable way to enable cooperative perception. For these reasons, we focus on an infrastructure-less approach, where vehicles coordinate directly amongst themselves in a distributed fashion to determine which vehicle shares which object’s point cloud.

**Contributions.** In this paper, we discuss the design and implementation of AUTOCast, a system that enables scalable infrastructure-less cooperative perception. Beyond extracting object points, AUTOCast uses several techniques to maximize the utility of information shared over the wireless channel. It determines *visibility* and *relevance* when deciding whether to transmit. For example, if car **A** wants to transmit roadway objects to car **B**, some of those objects may already be visible to **B**; **A** need only transmit objects *occluded* at **B**.

<sup>2</sup>Dedicated short range communication, a V2V standard.

<sup>3</sup>5GAA [1] extensions for vehicles potentially reach higher rates.

<sup>4</sup>LiDAR beams are radial, points are denser for nearby objects than for faraway ones. Objects faraway have smaller point cloud as well.

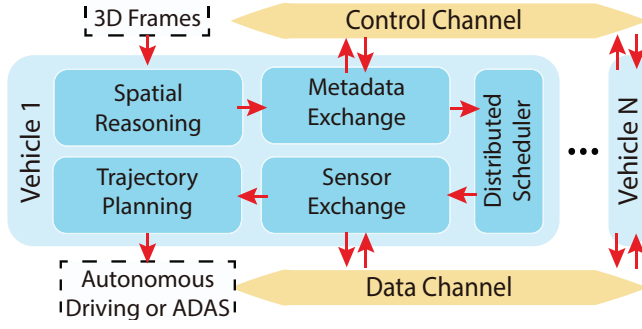


FIGURE 2: AUTOCAST system architecture

Moreover, **B** may not need all occluded objects, since some of them may not be *relevant* to its driving decision. AUTOCAST uses these criteria to determine which objects should be transmitted when.

To this end, AUTOCAST makes three contributions.

- A suite of fast *spatial reasoning* algorithms that analyzes point clouds to determine visibility and relevance.
- An efficient scheduling algorithm and a near-optimal heuristic that prioritizes safety critical transmissions.
- A planner that fuses shared points, estimates object motion, and finds collision free trajectories in a receding horizon.

We implement AUTOCAST in Carla [31], a photo-realistic autonomous driving simulator. Our evaluation results (§5) show that AUTOCAST can reduce all 100% hazardous situations (crashes, deadlocks) caused by occlusion in single-vehicle based perception, significantly reduce near-miss cases by providing early situational awareness and increasing reaction time. AUTOCAST's scheduler prioritizes most relevant information, scales gracefully up to 40 vehicles within sharing range, increasing visibility into safety critical objects by 2-4x of the time, 2-8x in visible size, and avoiding all collisions that occur using alternative baselines. We have also implemented a DSRC-radio based prototype for coordinated transmissions. Our experiments validate the DSRC channel capacity to meet this transmission schedule on time. Finally, we have optimized the end-to-end pipeline to operate at >10 fps.

## 2 AUTOCAST Architecture

AUTOCAST's end-to-end architecture (Figure 2) contains a *control-plane* (§3) that exchanges beacons and makes transmission scheduling decisions, and a *data-plane* (§4) that processes, transmits, and uses point clouds to make trajectory planning decisions. This decoupling of data and control ensures that bandwidth intensive point cloud data is directly transmitted between vehicles with *minimum delay for real time decisions*, while at the same time the control plane is able to make near optimal scheduling decisions.

**The control plane.** Two subcomponents constitute the control plane. The *metadata exchange* component (§3.1) implements a protocol to exchange metadata (needed for scheduling, obtained from the data plane) among vehicles. The *scheduler* (§3.2) uses this information to compute a transmission schedule, executed by the data plane.

**The data plane.** Two subcomponents constitute the data plane. *Spatial reasoning* (§4.1) extracts moving objects from LiDAR sensors. For each object, it determines which vehicles cannot see this object

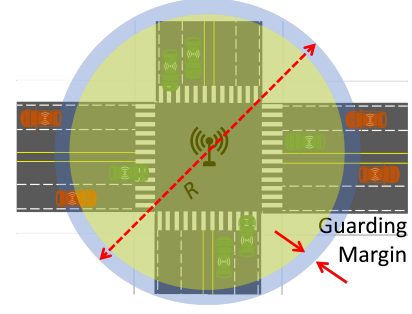


FIGURE 3: Scheduler domain (idealized): green vehicles are selected participants; red vehicles are excluded. A guarding margin is applied to account for vehicle movement during the decision interval.

and to which subset of them the object would be relevant; those are the vehicles to whom this object should potentially be sent. Each vehicle runs a *trajectory planning* component (§4.2), which fuses the received objects to adapt its current trajectory.

## 3 Control Plane

We now describe metadata exchange and scheduling components (Figure 2). Both the control plane and the data plane assume that each vehicle is able to accurately position itself using a 3-D map [19] and Simultaneous Localization and Mapping (SLAM [80, 81]), so that each vehicle knows its own position precisely at all times. This kind of positioning technology is mature and has been widely deployed.

### 3.1 Metadata Exchange

**Deployment setting.** Vehicles exchange metadata between themselves. Because the specifics of the metadata exchange can depend upon the relationship between radio range and road geometry, we describe metadata exchange for a concrete deployment setting, an *intersection*. Intersections are also where cooperative perception can help most [42], because they are among the most hazardous parts of the road network.

**Control messages.** AUTOCAST participants periodically broadcast *control messages* every  $T_d$ , the timescale at which the scheduler makes decisions (§1). These control messages are highly likely to reach each vehicle that is close to or at the intersection (Figure 3). This is because lane widths are on the order of 3-4 m [18, 77], so intersections of major streets with 3 lanes in each direction can be on the order of 30 m×30 m. On the other hand, the nominal radio range of LTE-direct, denoted as  $R$ , is over 170 m (urban non-line-of-sight) [5, 67], so even vehicles far away from the intersection can hear these messages. These control messages, or beacons, inform vehicles of their neighbors location, so that the scheduler can identify all those vehicles within a circle of radius<sup>5</sup> ( $R/2$ ); these are the vehicles that can plausibly hear each other and they constitute the scheduler's *domain* (Figure 3)<sup>6</sup>. We discuss this, together with how vehicles multiplex their transmissions to execute the schedule (§3.2).

<sup>5</sup>AUTOCAST associates a guard band to allow for vehicle movement. This guard band can be calculated from the posted speed limits: at 40 mph, a vehicle can move  $\sim 1.8$  m in one  $T_d$  (100 ms) interval.

<sup>6</sup>In practice, the radio range may be irregular. A smaller domain radius (e.g., 100 m, much less than maximum radio range  $R$ ) can be chosen to ensure all participants can plausibly hear each other.

**Information exchanged in control messages.** In AUTOCAST, participants exchange two types of information in these messages. Standardization efforts have defined V2V messaging formats that exchange similar *cooperative awareness messages* [35]; we have left it to future work to design a standard-compliant message exchange.

**Trajectory.** Each vehicle transmits its current *trajectory* to other participants; the vehicle's planner (§4.2) generates and updates the trajectory every  $T_d$ . A trajectory (denoted by  $t_i$  for vehicle  $i$ ) consists of a series of *waypoints* and their associated *timestamps*. Each waypoint indicates the position a vehicle expects to be at the corresponding timestamp. To limit control overhead, AUTOCAST down-samples trajectories into connected line segments for sharing. The first waypoint in the trajectory is the current *vehicle pose*.

**The object map.** Using its 3D sensor, each vehicle can extract point clouds of *roadway objects*; these are stationary or moving objects (vehicles, pedestrians, cyclists) on the road surface. Denote by  $o_{i,k}$  the  $k$ -th object in vehicle  $i$ 's view. Now, vehicle  $i$  receives broadcasted trajectories from other vehicles. Using spatial reasoning techniques described in §4.1, vehicle  $i$  computes the following two quantities for each  $o_{i,k}$  in its view: (1)  $v_{(i,k),j}$  is a boolean value that indicates whether  $o_{i,k}$  is visible to vehicle  $j$ . (2)  $r_{(i,k),j}$  is a value that indicates whether  $o_{i,k}$  is *relevant* to  $j$ 's current trajectory  $t_j$ . We make the notion of relevance precise in §4.1. Vehicle  $i$  then broadcasts an *object map* that contains: (a) an ID for each object  $o_{i,k}$ , (b) the size of  $o_{i,k}$  in bytes, (c)  $v_{(i,k),j}$ , and (d)  $r_{(i,k),j}$ . In §3.2, we explain how the scheduler uses these values to compute a transmission schedule.

**Loss compensation.** Control messages can be lost; In case of a loss, we reuse the trajectories from the most recent control message (evicting waypoints up to the current timestamp). AUTOCAST also *extrapolates* objects' current locations based on latest locations and timestamps. With extrapolated location, even when there is a packet loss, AUTOCAST recalculates the visibility ( $v_{(i,k),j}$ ) and relevance ( $r_{(i,k),j}$ ) metrics to update the object map.

**Multiple scheduler domains.** Figure 3 shows a scheduler domain at an intersection. Each intersection may have a distinct scheduler domain, and scheduler domains may exist along road segments as well. To accommodate these, scheduler domains can be predefined in high-definition maps. Using SLAM, vehicles can easily figure out which domain they belong to. To avoid inter-domain interference, neighboring domains can use different channels. Because the road network naturally isolates the map into blocks, we only need to assign scheduler domains with alternating channels along the length of the road, while each domain covers the full width of the road. Autonomous vehicles usually have multiple V2X antennas [10]. When they are at the border of two clusters, they can participate in both clusters at the same time and handover from one to the other following their moving direction. There is a large body of literature [20, 21, 34, 62, 73] enabling distributed clustering in the context of mobile ad hoc networks. In this work, we focus on the novel aspect of the system design, and leave the exploration of dynamic clustering to future work.

### 3.2 AUTOCAST Scheduler

In this section, we discuss AUTOCAST's scheduler. Depending on the network bandwidth and the number and size of objects  $o_{i,k}$  relevant to other vehicles, it may not be feasible to transmit all

relevant objects before the next decision interval  $T_d$ . The scheduler decides which objects to transmit at every decision interval and in what order. For example, if an object is likely to cause an imminent collision, it must have a higher priority in the transmission schedule. The underlying PHY layer may be able to transmit multiple PHY frames<sup>7</sup> during one decision interval; the scheduler must decide which objects to transmit in which frames. Each vehicle computes the schedule using the common list of control messages in its domain (Figure 3); each vehicle then broadcasts the specific object in the assigned PHY frame.

**Preliminaries: notation and PHY layer.** Let  $C = \{1, \dots, C\}$  be the set of vehicles and  $\mathcal{K} = \{1, \dots, K\}$  be the set of objects across all vehicles. Let  $x_{i,k}^n = \{0, 1\}$  be a decision variable indicating whether vehicle  $i$  transmits its object  $k$  at frame  $n$  (of  $N$  frames in total),  $S_f^n$  be the size of frame  $n$ , and  $T^n$  be the duration of frame  $n$ , thus  $\sum_n T^n \leq T^d$ , and  $T^n = S_f^n/B$  where  $B$  is the bandwidth.

$T^n$  and  $B$  depend on the PHY technology. There are two technologies available today, DSRC and LTE-V (a variant of LTE-direct).  $B$  varies between 5–10 Mbps and  $T^n$  varies between 10 and 100 ms in current standards. DSRC is based on TDMA while LTE-V offers both an OFDM option and a TDMA option. In case of OFDM, a frame multiplexes transmissions from multiple vehicles similarly to uplink frames in cellular networks. In case of TDMA, a frame consists of concatenated (in time) transmissions from multiple vehicles.

We assume the PHY layer uses QPSK as per common practice in vehicle communication systems due to the challenging channel [14, 15]. Thus, the system can deliver  $L = B \times \log(1 + \gamma_{QPSK})$  bits per unit time, where  $\gamma_{QPSK}$  is the SINR value required by QPSK. We model the PHY layer by the probability of successful delivery of an object between two vehicles. We define a  $C \times C$  channel matrix comprising of these probabilities as follows,  $\mathbb{P} = [p_{i,j}, i, j = 1 \dots C]$ , where  $p_{i,j} \in [0, 1]$  indicates the probability of delivery from vehicle  $i$  to  $j$ ,  $\forall i, j \in C$ , ( $p_{i,j}$  are assumed independent).

**Problem formulation: Markov Decision Process.** Because a decision interval may have multiple frames, we formulate the scheduling problem as a Markov Decision Process (MDP) that optimizes the schedule across all frames.

**State.** Let  $h_{(i,k),j}^n$  indicate whether vehicle  $j$  has received object  $k$  from vehicle  $i$  by frame  $n$ , and  $q_j^n = \{h_{(i,k),j}^n, \forall i, k\}$ . We define the state of the system at frame  $n$  by  $S^n = \{q_1^n, q_2^n, \dots, q_C^n\}$ ,  $S^n \in \mathcal{S}$ , with  $\mathcal{S}$  denoting the state space. Since the state changes for each frame,  $n$  represents the discrete time steps over which the MDP operates.

**Action.** Let  $s_{i,k}$  denote the size of object  $o_{i,k}$ ,  $A^n = \{x_{i,k}^n = \{0, 1\}, \forall i, k \mid \sum_{i \in C} \sum_{k \in \mathcal{K}} s_{i,k} \times x_{i,k}^n \leq S_f^n\}$  denote the action taken at time step  $n$ , where  $A^n \in \mathcal{A}$ , the action space.

**Reward function.** To maximize the total rewards the system needs to carefully decide the action ( $A^n$ ) based on the current state ( $S^n$ ). When the action is decided, the reward follows:

$$R^n = \sum_{j \in C} \sum_{i \in C} \sum_{k \in \mathcal{K}} x_{i,k}^n \times (h_{(i,k),j}^{n+1} - h_{(i,k),j}^n) \times y_{(i,k),j}^n, \quad (1)$$

<sup>7</sup>The term "frame" used in this section refer to the time steps in a decision interval  $T_d$ , which is different from the lidar frame in §4



where  $y_{(i,k),j}^n$  is the reward when object  $k$  is transmitted from vehicle  $i$  to vehicle  $j$ , which we define by

$$y_{(i,k),j}^n = (1 - v_{(i,k),j}) \times r_{(i,k),j}.$$

The rationale for this definition is that there is a reward if vehicle  $j$  receives object  $k$  by vehicle  $i$  if object  $k$  is invisible and relevant to vehicle  $j$ , see §3.1.

**Transition probability.** We compute the transmission probability from one state to another based on action  $A^n$  as follows:

$$P_{S^n, S^{n+1}}^{A^n} = \prod_{i: x_{i,k}^n = 1} \left( \prod_{j \in \mathcal{V}^1} p_{i,j} \times \prod_{j \in \mathcal{V}^0} (1 - p_{i,j}) \right), \quad (2)$$

where  $\mathcal{V}^1 = \{j | (h_{(i,k),j}^{n+1} = 1, h_{(i,k),j}^n = 0, x_{i,k}^n = 1)\}$  corresponds to all vehicles  $j$  which successfully received the scheduled object during time step  $n+1$ , whereas  $\mathcal{V}^0 = \{j \in C | (h_{(i,k),j}^{n+1} = 0, h_{(i,k),j}^n = 0, x_{i,k}^n = 1)\}$  corresponds to vehicles which lost the scheduled object during that frame.

**Markov Decision Process.** We define a finite-horizon MDP by the tuple  $\mathcal{M}(\mathcal{S}, \mathcal{A}, P_{S^n, S^{n+1}}^{A^n})$ . To solve the MDP, we first define a policy  $\pi$  to be a mapping from states to actions and seek to find the optimal policy which maximizes the (expected, discounted) sum of the rewards occurring from the selected actions over a (potentially infinite) time horizon.

To find the optimal policy a recursive approach is used, which updates (i) policy decisions  $\pi(S^n)$  at each frame  $n$  and (ii) the value function  $\mathcal{U}^\pi(S^n)$  which keeps track of the sum of the rewards if policy  $\pi$  is followed from state  $S^n$ . Specifically, the corresponding recursive formula is given by:

$$\mathcal{U}^\pi(S^n) = \sum_{j \in \mathcal{V}^{n+1}} P_{S^n, S^{n+1}}^\pi \cdot (R^n + \mathcal{U}^\pi(S^{n+1})). \quad (3)$$

The rationale behind this equation is that an optimal policy can be constructed by going backwards in time: we first construct an optimal policy for the tail subproblem corresponding to the last step at time step  $n = N$ , then apply the optimal policy to the tail subproblem corresponding to the last two steps at time step  $n = N-1$ , and continue in this manner until an optimal policy for the entire problem is formed.

One may use dynamic programming (DP) to find the optimal policy  $\pi^*(S^n) \in \mathcal{A}$  which maximizes the value function in Equation 3 using the Bellman equation [24]. However, the time complexity grows exponentially [63] with the number of states. This motivates us to seek more scalable approaches.

**Scheduling algorithms.** We start by defining the weight of an object at frame  $n$ , denoted by  $H_{i,k}^n$ , to be the total rewards gained by the system if the object is successfully delivered to all interested vehicles:

$$H_{(i,k)}^n = \sum_{j \in C, i \neq j} y_{(i,k),j}^n \times (1 - h_{(i,k),j}^{n-1}). \quad (4)$$

Note that the term  $(1 - h_{(i,k),j}^{n-1})$  indicates the object has not been received during previous frames.

Consider a bipartite graph (Figure 4) which has destination vehicles on the left and objects on vehicles on the right. The weight of each edge between a vehicle and an object corresponds to the reward if this vehicle receives that object. Then,  $H_{i,k}^n$  can be computed by adding the weights of the edges connecting to the  $(i, k)$  node.

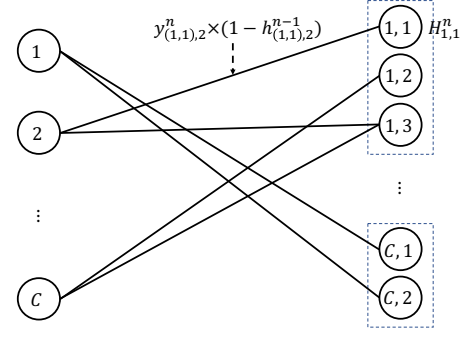


FIGURE 4: An illustration of  $H_{i,k}^n$ .

**Greedy max-weight scheduler.** Motivated by this representation, we may use greedy solutions to the maximum weight matching problem of a bipartite graph [76] to quickly find a good solution. Specifically, at every frame/time step, the scheduler may select the transmission pairs based on decreasing order of the  $H_{i,k}^n$  value, leading to the highest possible total weight/reward among the available transmissions for each frame, until there is nothing to deliver or the decision interval is over (i.e.,  $n = N$ ).

However, the above weight does not take the size of an object,  $s_{i,k}$ , into account. Therefore, the scheduler may schedule an object that has a large weight but occupies a large portion of the frame, as opposed to scheduling a large number of smaller objects whose sum of weights might be larger than the weight of the single large object. One way to address this is to divide the weight of an object by its size, and use the modified weight,  $H_{i,k}^n/s_{i,k}$  instead, corresponding to a *normalized reward*  $y_{(i,k),j}^n/s_{i,k}$ , over the size of the object.

---

#### Algorithm 1 Greedy Max-Weight Scheduler

---

**Input:**  $y_{(i,k),j}^n$ ,  $h_{(i,k),j}^n$ ,  $s_{i,k}$ , and  $S_f^n$

**Output:**  $\pi(S^n) = \{x_{i,k}^n\}$

1. **for**  $n \in \{1, \dots, N\}$  **do**
  2.   Calculate  $H_{i,k}^n$  from Eq. (4).
  3.   **while**  $\sum_{i \in C} \sum_{k \in \mathcal{K}} s_{i,k} \times x_{i,k}^n \leq S_f^n$  **do**
  4.     Select a TX pair  $(i, k)$  with the largest  $H_{i,k}^n/s_{i,k}$ .
  5.     Set  $x_{i,k}^n = 1$ .
  6.   **end while**
  7.   Update  $h_{(i,k),j}^n$  based on the vehicle environment.
  8. **end for**
- 

We summarize in pseudo code the proposed greedy Max-Weight algorithm (Algorithm 1). The time complexity of the scheduler can be easily shown to equal  $O(NCK \log(CK))$ . CPU experiments show that the Greedy Max-Weight scheduler runs fast and achieves near optimal performance for the scenarios that we have studied (§5).

**FPTAS-based scheduler.** We also propose to use a well known fully-polynomial time approximation scheme (FPTAS) [32] to solve the selection problem at every time step. We first introduce a dynamic programming framework which solves the following equation:

$$DP(C \times \mathcal{K}, S_f^n) = \max \{ DP(C \times \mathcal{K} \setminus (i, k), S_f^n), H_{i,k}^n + DP(C \times \mathcal{K} \setminus (i, k), S_f^n - s_{i,k}) \}. \quad (5)$$

We then formulate the scheduling problem at each time step as a binary Knapsack problem, and solve it using FPTAS. While more efficient than dynamic programming, FPTAS still has high computational complexity (see §5).

**Starvation compensation.** When network capacity is insufficient to transmit all objects, some objects may “starve” (not be transmitted). To avoid this, AUTOCASD discards stale updates (point clouds), increases such objects’ weight to maximize the likelihood that new updates (if still relevant with positive rewards) are transmitted in a future decision interval. Let  $m$  represent the number of decision intervals over which an object has *not* been transmitted even though it is still occluded and relevant. We replace in Algorithm 1 the value of  $H_{i,k}^n/s_{i,k}$  with  $\frac{H_{i,k}^n}{s_{i,k}/\sigma_{s_{i,k}}} \times m/\sigma_m^2$ , where  $\sigma^2$  denotes the variance and is used to make sure the contributions of the size and the starvation effects are normalized.

**Other details.** When a schedule is decided, the cooperative execution of that schedule among vehicles has differences depending on which V2V technology is used. DSRC uses TDMA among vehicles. LTE-V may use TDMA (mode 4) or SC-FDMA (mode 3), an OFDM variant, where frequency-time slots can be assigned to vehicles based on the schedule [15]. AUTOCASD can employ DSRC or any LTE-V mode; in §5, we use real DSRC radios to demonstrate the scheduled transmissions, and have left LTE-V integration to future work.

## 4 Data Plane

Autonomous vehicles use 3D sensors for *perception*, and a *planning algorithm* to determine the vehicle’s trajectory. AUTOCASD proposes to extend today’s autonomous driving with *cooperative perception*. Its data plane achieves cooperative perception using *spatial reasoning* algorithms that generate object maps (§3.1), and a *planner* that relies on cooperative perception to improve driving safety.

### 4.1 Spatial Reasoning

This component processes each frame of the LiDAR output and generates the object maps. Specifically, for each object  $k$  in vehicle  $i$ ’s view, spatial reasoning determines the *visibility*  $v_{(i,k),j}$  of that object with respect to another vehicle  $j$ , and the *relevance*  $r_{(i,k),j}$  of that object to that vehicle (§3.1). To do this, it must (a) detect roadway objects within its view, (b) assess their geometric and temporal relationships.

**Extracting roadway objects.** Several deep learning networks exist [51, 74, 79, 83] that can detect objects in LiDAR frames. However, these can be slow, require significant compute resources, are sometimes inaccurate, and generate information (e.g., identify object classes and bounding boxes) that AUTOCASD does not need. For AUTOCASD, we simply need point clouds of stationary or moving objects on the road.

To extract these roadway objects, we voxelize [51, 55] the point cloud by imposing a fine 2-D occupancy grid from the birds-eye-view perspective of the LiDAR point cloud (Figure 5). More precisely, each 2-D grid element is a rectangular tube extending vertically on the z-axis. Each point in the LiDAR frame falls into exactly one grid element. Each grid element may contain: (a) no points, (b) only points on the ground, or (c) points above ground. AUTOCASD can determine if a point lies on the ground or above the ground because it knows the coordinates of the point, and the height of the

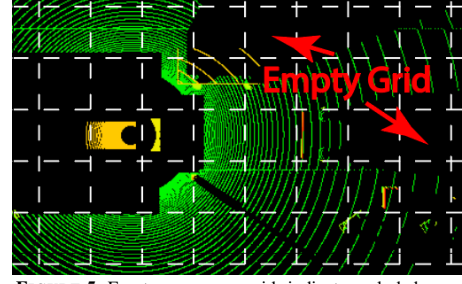


FIGURE 5: Empty occupancy grids indicate occluded area.

LiDAR above the ground. Furthermore, AUTOCASD assumes that each 2-D grid is labeled as either on the roadway surface or not. This information can be obtained by running a segmentation algorithm for drivable space detection on the 3-D map [59, 75]. Thus, all points in a 2-D grid element of type (c) which are above the ground constitute points belonging to a roadway object. The object itself consists of all contiguous type (c) grid elements (as we discuss later in §5, we use sub-meter grid dimensions so it is unlikely that points belonging to two different vehicles would fall into the same grid). Objects detected by different vehicles with the same grids (after perspective transformation at the receiver (§4.2)) are identified as the same unique object.

**Visibility determination.** With extracted objects, to determine  $v_{(i,k),j}$ , vehicle  $i$  simply traces a ray from vehicle  $j$ ’s current position around every object  $o_{i,k'}$  in its own view ( $o_{i,k'}$  includes the vehicle  $i$  itself, if visible). If  $o_{i,k}$  falls into the shadow of  $o_{i,k'}$ , then the latter object occludes the former and  $v_{(i,k),j}$  is false. If no such  $o_{i,k'}$  exists, and if  $o_{i,k}$  is within the LiDAR range of  $j$ , then  $v_{(i,k),j}$  is true.

**Relevance determination.** The intuition behind relevance determination is that some objects may not be relevant to other vehicles, even if invisible to those vehicles. For instance, if a vehicle is turning right at an intersection, it is unlikely to need information about vehicles driving straight on the opposite lane. In AUTOCASD’s implementation, an object is relevant to another vehicle if the trajectories of those two could potentially collide at some point in the future.

More precisely,  $r_{(i,k),j}$  is a value that assesses whether vehicle  $j$ ’s trajectory can collide with  $o_{i,k}$ . Vehicle  $i$  gets  $j$ ’s trajectory from control messages. It obtains  $o_{i,k}$ ’s trajectory by estimating this objects’ heading and velocity continuously over successive frames. By extrapolating these trajectories, AUTOCASD can determine if the two trajectories collide at some point in time. Given this, one can define  $r_{(i,k),j}$  in two ways: (a) as a boolean value that is true when  $j$  can collide with  $o_{i,k}$ , or (b) as the reciprocal of the time to collision (a value between 0 and 1, assuming time is in milliseconds). The intuition for the latter choice is clear: objects that  $j$  is likely to encounter sooner are more relevant<sup>8</sup>.

**Loss compensation.** Similar to handling control message losses (§3.1), data packet losses are also compensated by extrapolation. Given a loss of a particular object, AUTOCASD calculates the center of the object point cloud using the velocity estimated from the location and timestamp of previous receptions. Then, it translates the last received point cloud to the newly estimated location.

<sup>8</sup>At a large scale, boolean and reciprocal definitions perform similarly (Equation 1). For simplicity, we used the boolean definition in our evaluation.

## 4.2 Trajectory Planning

Autonomous vehicles use sensor inputs to make driving decisions. These driving decisions occur at three different scales: *route planning* occurs at the granularity of a trip, *path and trajectory planning* occurs at the granularity of a road segment (a few tens to hundreds of meters), and *low-level control* ensures that the vehicle follows the planned trajectory by effecting steering and speed control. In AUTOCAST, vehicles must make these decisions, by incorporating received point-clouds into their own LiDAR output.

In this paper, in order to quantify the end benefits of cooperative perception, we develop a path and trajectory planning algorithm that incorporates objects received from other vehicles. The large, existing literature on this topic (see, for example, [43, 45, 52]) does not take cooperative perception into account. Recent research has recognized and incorporated partial visibility into trajectory planning [33, 68], but relies on training data and predictions of the geometry of the invisible area. In contrast, we develop a planning algorithm using concrete cooperative perception for trajectory planning.

**Perspective transformation.** Before it can plan a trajectory, AUTOCAST must *re-position* the received point clouds into its own LiDAR output. It uses the 3-D map for this. The sending vehicle positions the point cloud in its own coordinate frame of reference. To re-position it to the receiver, let  $T_s$  be the transformation matrix from the sender's coordinate frame of reference to that of the 3-D map and  $T_r$  be the transformation matrix for the receiver. To transform a point  $V_s$  in the sender's view to a point  $V_r$  in the receiver's, AUTOCAST uses:  $V_r = T_r^{-1} * T_s * V_s$ . Having done this, it updates each occupancy grid (§4.1) with the received point cloud, then uses the occupancy grid to determine a *path* and then a *trajectory*.

**Path Planning.** This step determines a viable and safe path through drivable space that avoids all objects. It uses the occupancy grid defined above (§4.1) after augmenting it with received objects. To understand path viability, recall that each grid element can either have one or more points belonging to an object, or be *unoccupied*. Moreover, using the 3-D map, we can annotate whether a grid element belongs to *drivable* space or not, and also whether a vehicle can traverse the grid element in both directions or uni-directionally.

The input to path planning is a source grid element and a target grid element. The output of path planning is a *path* in the 2-d grid, where the width of the path is the width of the car, and every grid element that intersects with the path must (a) be unoccupied and (b) be drivable in the direction from the source to the target. AUTOCAST uses A\* heuristic search [44] to determine a valid path. We constrain the search so that the resulting path is *smooth*: i.e., it does not have sharp turns that could not be safely executed at the current speed.

**Trajectory Planning and Collision Avoidance.** On the resulting path, AUTOCAST picks equally spaced *waypoints*; a *trajectory* is a collection of waypoints and associated times at which the vehicle reaches those waypoints. To determine those times, the trajectory planner must determine a *collision-free* trajectory; when the vehicle is at a particular waypoint, all other vehicles must be far enough from that waypoint. To determine this, AUTOCAST uses the estimated trajectory of received objects, as well as estimates of the trajectory of vehicles within its own sensor's view. AUTOCAST also calculates the time of arrival to and departure from this waypoint based on estimated speed and vehicle dimension. When a predicted collision

is far enough, AUTOCAST follows the planned trajectory until within stopping distance (based on current speed and brake deceleration) of that waypoint of collision.

## 5 Performance Evaluation

In this section, we first evaluate AUTOCAST end-to-end: we show that cooperative perception can improve driving safety on three autonomous driving benchmarks (§5.2). We break down the results by traffic density to discuss scalability (§5.3), and then detail the results of each scenario (§5.4). Next, we evaluate the latency of each processing module (§5.5), and validate transmissions using real V2V radios (§5.6). We conclude with micro-benchmark evaluations on the scheduling details (§5.7).

### 5.1 Methodology

**The Carla simulator.** Carla [31], a photo-realistic simulation platform, uses a game engine to simulate the behavior of realistic environments, and contains built in models of freeways, suburban roads, and downtown streets. Users can create vehicles that traverse these environments and attach advanced sensors such as LiDAR, Camera, Depth Sensor to them. As these vehicles move through the environment, Carla simulates environment capture using these advanced sensors. Users can design planning and control algorithms using the captured environment to validate autonomous driving.

**Implementation.** We have implemented the scheduler, spatial reasoning, and trajectory planning in Carla. The total AUTOCAST implementation is 27,124 lines of code. In addition, to configure the scenarios, we have developed on top the Carla autonomous driving challenge [6] evaluation code. In our implementation, all vehicles use Carla's default longitudinal and lateral PID controller as the lower-level control to steer the vehicle along the planned trajectory.

To simulate metadata exchange between vehicles, we have incorporated V2V. Specifically, our implementation models LTE-Direct QPSK with 10 MHz bandwidth [39], which translates to a peak rate of  $\sim 7.2$  Mbps. We implement LTE-Direct TDMA Mode 4 [39] and simulate V2V channel loss in all scenarios using models described in 3GPP standards [14, 15].

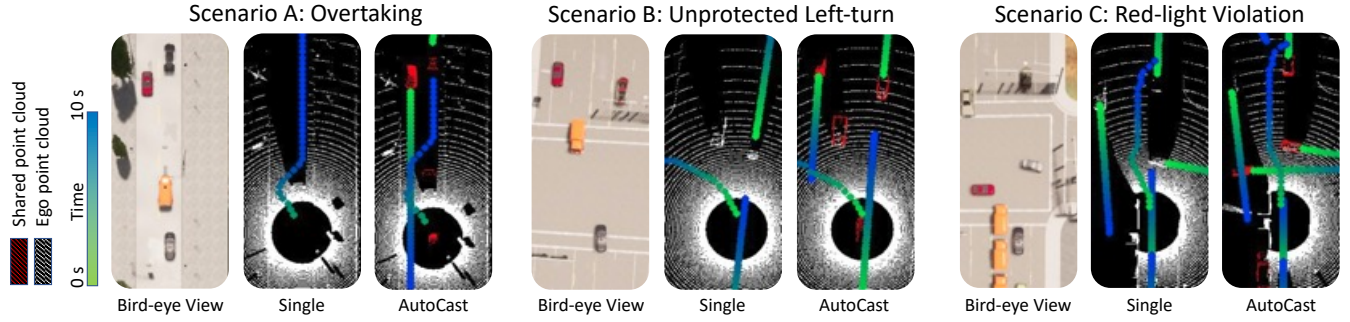
**End-to-end evaluation scenarios.** To demonstrate the benefits of AUTOCAST end-to-end, we have implemented three scenarios (Figure 6) from the US National Highway Transportation Safety Administration (NHTSA) Pre-crash typology [60]. In these, occlusions can impact driving decisions.

**Overtaking** A stopped truck on a single-lane road forces a car to move to the lane with on-coming traffic. The truck occludes the car's view of the opposite lane.

**Unprotected left turn** A car and a truck wait to turn left in opposite directions at an intersection. The truck blocks the car's view oncoming traffic.

**Red-light violation** A truck waits to turn left at an intersection, and a car drives straight towards the intersection. Another car jumps the red-light in the perpendicular direction; the violator is occluded by the truck.

**Experiments with real radios.** To demonstrate that an implementation of AUTOCAST can plausibly work over real radios, we run AUTOCAST on a small-scale testbed using three iSmartWays DSRC radios [9]. In these experiments, we record the trace data from all



**FIGURE 6:** End-to-end evaluation scenarios: overtaking, unprotected left-turn, and red-light violation. A planner on the *ego* vehicle (gray, bottom of the bird-eye view) finds a trajectory to navigate through each scenario without collision. The gradient trajectory color (green to blue) indicates a temporal horizon (closer to farther). The LiDAR views show the perception results using either non-sharing baseline (*Single*), or AUTOCAS. The red points in the LiDAR view are shared points, while the white ones are from the ego vehicle itself. In each scenario, a *passive* (without communication capability) *collider* vehicle (red), occluded by a truck (orange) and thus invisible from the ego's *Single* view, may cause a hazardous situation. AUTOCAS makes the ego vehicle aware of the collider so the ego can react early to avoid a collision.

scenarios; for each frame (every 100ms), the trace includes all exchanged metadata, the computed schedule, and the object point clouds. We then playback the trace over DSRC radios to validate if the scheduled transmissions complete in time.

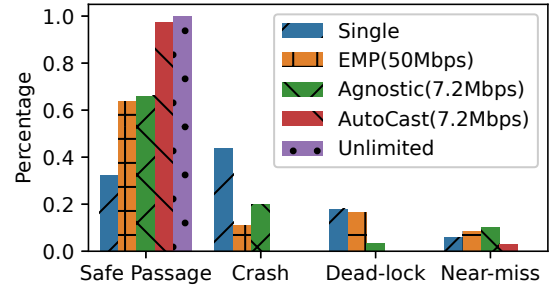
**Baselines.** We compare AUTOCAS against an approach in which each car makes trajectory planning decisions based on its own sensor alone (called *Single*), and one in which cars within range deliver objects in a round-robin fashion (called *Agnostic*). We also implemented *EMP* [82], an edge-assisted cooperative perception scheme, as another baseline. Because *EMP* is a V2I solution instead of V2V, for fair comparisons, we assign *EMP* a total bandwidth of 50 Mbps, but use 7.2 Mbps for *Agnostic* and AUTOCAS. Finally, we also compare against an idealized baseline, *Unlimited*, in which all points can be transmitted without bandwidth constraints. In quantifying the efficacy of our greedy scheduler, we also compare it with 1) *FPTAS*, 2) Optimal using dynamic programming.

**Metrics.** We use several metrics to evaluate AUTOCAS. In end-to-end experiments, we quantify scenarios *outcomes* (e.g., a crash, or a near miss), the *reaction* time (between when a vehicle detects a potential collision and the time needed for it to avoid the collision), and the *closest distance* between two vehicles at any point during the scenario. To analyze the scalability, we compare the *collider* *visibility* in terms of the number of visible frames and shared points under different schemes at different traffic densities. To evaluate the scheduler's efficacy, we quantify *rewards*, *time complexity*, and *scheduled delay* of objects with different rewards.

## 5.2 End-to-end Scenario Evaluation

**Goal.** The NHTSA pre-crash typology defines a set of challenging scenarios. In this section we seek to understand whether cooperative perception can result in safer driving outcomes than a system without this capability. We evaluate these scenarios in Carla: for each scenario, we explore different points of the scenario parameter space (described below), and record the metrics described above.

**Terminology and Experiment Setting.** In each of our scenarios (Figure 6), there are three entities: the gray sedan is the *ego* vehicle on which AUTOCAS runs, the red sedan is a *passive collider* which cannot communicate and only uses its own sensors to plan its trajectory, and the orange truck is an *occluder*. In each scenario, the paths



**FIGURE 7:** Scenario Outcome

between the ego and the collider intersect. We set up their speeds such that their *trajectories* almost collide (i.e., both vehicles would come very close to each other if both did not see each other at all).

Specifically, we generate several *configurations* as follows. We set the *base speed* of the collider to 3 different values (20 km/h, 30 km/h and 40 km/h). At a given base speed, the collider's trajectory would (in the absence of avoidance) intersect with that of the ego. A second dimension of the configuration is an *intersection delta*; ranging from -2 s to +2 s (with steps of 0.25 s), a value of  $\delta$  means the collider actually arrives at the intersection point  $\delta$  s before (or after) the ego vehicle. This latter parameter controls how closely the two cars approach each other. This gives us a total of 24 different configurations for each scenario. For each configuration, we also vary the number of vehicles (from 5 to 40 within range R (§3.1)) to evaluate the performance at different scales.

We present three sets of results. First, we present the end-to-end results for all scenarios that highlights AUTOCAS's performance against *Single*, *EMP*, *Agnostic*, and *Unlimited*. Then, we break it down by traffic density to highlight the scalability (§5.3) against the baselines. Finally, to illustrate subtleties in AUTOCAS's design, we break down results by scenario (§5.4), in a sparser traffic setting.

**Outcomes.** With this setting, there are four possible outcome: *safe passage*, *near-miss*, *crash*, and *deadlock*. A near miss occurs when the ego and collider pass within 2 m of each other. In deadlock, which occurs only in Overtaking, both vehicles stop without colliding but neither can make forward progress. This situation is not inherently unsafe, it does represent an undesirable driving outcome where



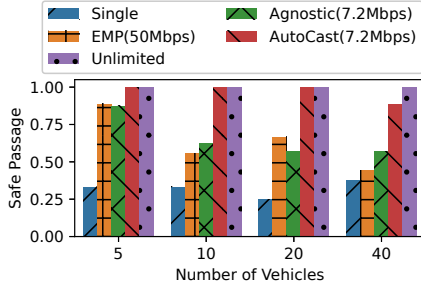


FIGURE 8: Safe passage vs. traffic density

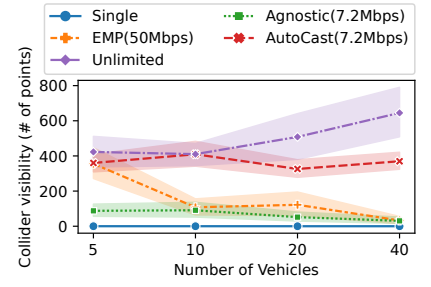
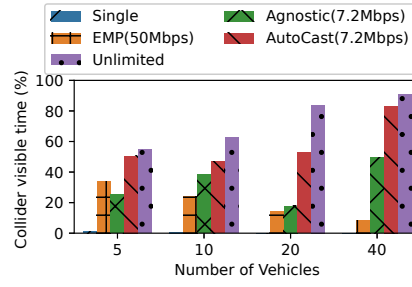


FIGURE 9: Collider visibility: visible time and visible size vs. traffic density

participants must coordinate to resolve the deadlock. Beyond these outcomes, we are also interested in quantifying the closest distance between vehicles during any point in the simulation, and the reaction time (the time between when the ego is aware of the collider to the last possible moment before it can start braking).

**Results: Dense Traffic.** Our first set of results demonstrate the efficacy of AUTOCAST in settings that it was designed for: highly dense settings with a large number of traffic participants, where channel capacity precludes transmission of point clouds of all participants. In this experiment, we ran all three scenarios (overtaking, left turn, and red light violation), but varied the traffic density from 5 to 40 vehicles within range R (§3.1). For each configuration of each scenario, we additionally vary the collider speed from 20 to 40 km/h.

Figure 7 plots the fraction of outcomes for each of the five alternatives we consider (*Single*, *EMP*, *Agnostic*, *AUTOCAST*, and *Unlimited*). *AUTOCAST* ensures safe passage at all collider speeds in all scenarios. By contrast, *Single*, which does not use cooperative perception, incurs crashes about half the time, many deadlocks and some near misses. *EMP* [82] divides full point clouds into segments of a voronoi diagram. Each vehicle transmits the closest segment to an edge server, which then can forward to other vehicle recipients. Using an order of magnitude higher V2I communication bandwidth, *EMP* safely passes half of the traces, incurs about 20% crashes and 20% deadlocks, most of which happens at high density scenarios. *EMP* does not consider object relevance, or prioritize segments, so suffers from lower collider visibility (see scalability results in §5.3). *Agnostic*, which also does not prioritize transmissions, but benefits from object-based transmission, passes 20% more traces with lower V2V bandwidth. Similar to *EMP*, at higher density, *Agnostic* exhibits an undesirable outcome in about a third of scenario settings. *Unlimited* achieves 100% safe passage, which validates the robustness of our planner design across our scenario configurations and collider speed choices.

We discuss differences between *AUTOCAST*, *EMP* and *Agnostic* below. *AUTOCAST* outperforms *EMP* and *Agnostic* for two reasons: (a) it extracts objects and prioritizes on cheap transmissions based on visibility and relevance, and, (b) when data to be transmitted exceeds the channel capacity, and some objects have to be left out, *AUTOCAST* compensates by prioritizing these objects in the next decision interval. This ensures consistent and continuous updates for critical invisible objects. In *EMP* and *Agnostic*, due to delayed

or missing updates, the closest distance to the collider is often below 2 m, leaving the ego vehicle almost 0 reaction time.

Next, we show more details by analyzing why *AUTOCAST* can outperform other baselines at scale (§5.3), and breaking down the results by scenario (§5.4) for detailed analysis.

### 5.3 Scalability Results

We now show the scalability of *AUTOCAST* by breaking down the evaluation results (Figure 7) by traffic density.

**Safe passage at different traffic densities.** Figure 8 shows the percentage of safe passage of all scenarios with the number of vehicles varying from 5 to 40. The *Single* baseline shows a uniform performance across traffic densities: using ego vehicle’s LiDAR alone can only pass around 25% of the traces. It is interesting to see at very high density (40 vehicles), the rate is slightly higher because the road becomes crowded, forcing all participants, including the ego vehicle, to slow down and proceed with caution. In contrast, *EMP* and *Agnostic* achieve on average around 60% safe passages, but most success concentrate at low density. As the traffic density increases, it is harder to avoid collisions and near-misses. We discuss key insights into poor scalability of both approaches. Given an order of magnitude higher bandwidth for V2I, *EMP* is able to share almost all Voronoi segments [82] at low density to cover the entire area. However, as the number of vehicles increases, each vehicle is transmitting a smaller and smaller segment around it, where the point cloud is the densest. Therefore, the total number of points to share, covering the same area, increases and exceeds even the V2I bandwidth limit. Also, since *EMP* does not prioritize segments based on relevance to receivers, the ego vehicle is aware of the collider only when the particular segment, where the collider is in, gets transmitted. *Agnostic*, which isolates objects to reduce bandwidth requirement, but does not prioritize object transmission, incurs a few near misses at low densities. As density increases, *Agnostic* incurs more crashes and deadlocks; with 10 vehicles and above, the probability of collider being transmitted is very low; *Agnostic* exhibits an undesirable outcome in about half of scenario settings. Over all traffic densities, *AUTOCAST* is near-perfect: it maintains 100% safe passage to 20 vehicles; at the highest traffic density we have evaluated, it incurs a small number of near misses without any collisions. Using only

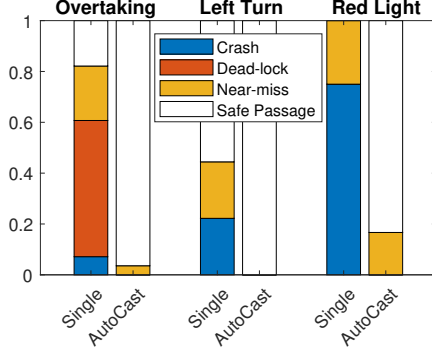


FIGURE 10: Per-scenario outcome at low-density

7.2 Mbps, AUTOCAST scales gracefully up to 40 vehicles in a distributed fashion, not depending on infrastructure support.

**Collider visibility.** To better understand the scalability results, we pick one intersection scenario (Scenario C: red-light violation) to evaluate and compare the *collider visibility* against different baselines. Figure 9 shows the percentage of frames where collider is visible (left) and the size of the shared point cloud of the collider (right). *Unlimited* depicts the maximum physical visibility of the collider from all vantage points in this particular scenario. Because both *Agnostic* and *EMP* does not prioritize transmissions, the chance of the collider being transmitted is highly dependent on the number of objects (or segments in *EMP*'s case) to be transmitted and their sizes: at low traffic density, *EMP* is likely to transmit all segments, *Agnostic* has a higher chance to transmit the collider among less other objects; at high density, *EMP* cannot transmit all segments, and *Agnostic* renders lower collider visible time as well. It is critical for the planner to have high and stable collider visibility, because how often and how big the point cloud of the collider gets transmitted directly determines the object detection accuracy, reaction time, the trajectory planning, the control decisions, and therefore the scenario outcome (see per-scenario analysis §5.4). AUTOCAST enables safe passage at high traffic density by providing consistent and stable (>80% of the time) visibility into occluded and relevant objects to avoid safety hazards.

## 5.4 Per-scenario Results

We now illustrate the benefit of cooperative perception in each of these three scenarios. We do this at low density, because the traffic dynamics are simpler and the results are easier to understand. For these results, we omit *Agnostic* and *EMP* because at low density the performance is comparable.

**Overtaking.** Figure 10 shows, for AUTOCAST and *Single*, a stacked bar that counts the number of outcomes of each type. Without cooperative perception, safe passage occurs only in 20% of the configurations. Crashes occur in about 5% of the configurations. These occur because the ego vehicle moves into the oncoming lane, but neither vehicle has enough time to stop. About 20% of the configurations result in near misses, and the remaining settings result in deadlocks. By contrast, AUTOCAST ensures safe passage in all configurations, but incurs near misses in about 5% of the configurations.

AUTOCAST's cooperative perception enables (Figure 11) much higher reaction times (average 13.3 seconds at 20 km/h), than without (average 0.31 s at 20 km/h, and zero at higher speeds), explaining

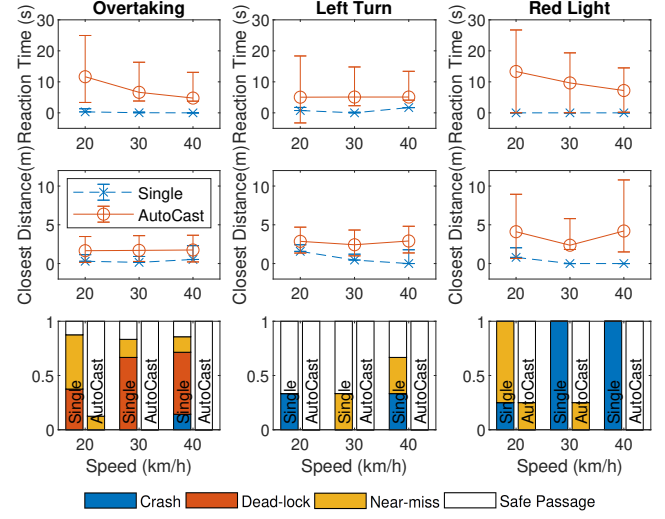


FIGURE 11: Reaction time, near-miss and crash details

its better performance. Finally, we have also investigated why AUTOCAST incurs some near misses (with a minimum closest distance of 1.74 meters) in this scenario. At low collider speeds, the ego is aware of the oncoming vehicle for 13 seconds and plans to stop the car as close and safe as possible to start lane change right after the collider passes; this represents a benign near-miss.

**Unprotected left-turn.** This scenario is more benign than overtaking, because the ego is obstructed to a lesser extent. Without AUTOCAST, Figure 10 shows that 16.7% of the configurations resulted in a crash and 16.7% in a near miss. For this scenario, AUTOCAST ensures safe passage in all configurations. Because this scenario is benign, reaction times are in general higher both with and without AUTOCAST. Without AUTOCAST, crashes occur when the collider arrives in the shadow of the truck as the ego vehicle starts to turn left. When the collider's speed is high, it cannot brake fast enough to avoid the collision or near miss. With AUTOCAST, reaction times and closest distances are generally quite high.

**Red-light violation.** For this scenario, *Single* always incurs an undesirable outcome (75% crashes, 25% near miss). The occlusion angle is so wide that there is no time for the ego vehicle to react whatsoever (0 sec in the third column of Figure 11). By contrast, AUTOCAST only incurs a few near-misses at low speeds: the controller is aware of the collider, but decides to leave little room to pass by.

## 5.5 Pipeline Micro-benchmark

AUTOCAST carefully optimizes the perception and planning pipelines to achieve around 80 ms end-to-end processing latency. Figure 12 shows the average processing time and its variance for each module over 6000 frames, 2000 frames for each scenario. Object extraction and path planning are the most compute intensive. We use the Minkowski Engine [29] for fast sparse quantization and Numba [11] to speed up python loop execution such as the extended A\* search and the isolated island detection. The execution time of each module may or may not be dependent on various factors. Quantization only depends on the number of lidar inputs. Creating the occupancy grid and extracting surrounding objects is depends on environment complexity, so their variances are high, but the

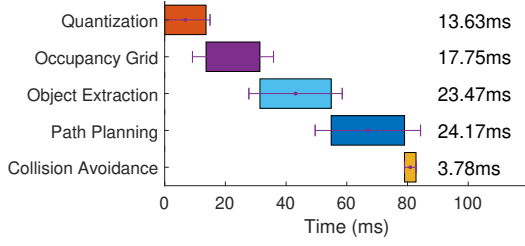


FIGURE 12: Pipeline micro benchmark

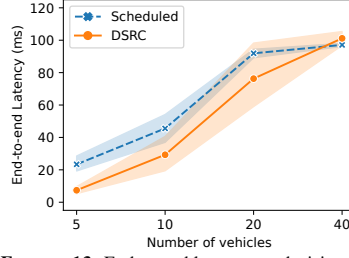


FIGURE 13: End-to-end latency per decision interval.

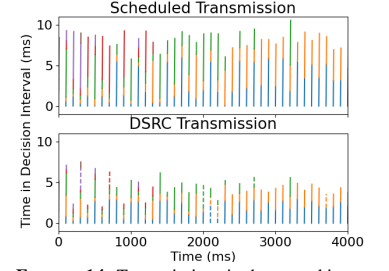


FIGURE 14: Transmissions in the overtaking scenario at low density

maximum variance is below 20%. Path planning takes on average 24.17 ms, but its results can be reused across frames with a fast waypoint check ( $\sim 10$  ms) across frames. The total latency is well below 100 ms, so AUTOCAST can process LiDAR at full frame rate (10 fps). To sense and react to the dynamics in the environment, sensor suite (LiDAR, Stereo Cameras, *etc.*) on autonomous vehicles are clocked at a minimum of 10 fps. Therefore it is critical to optimize the processing pipeline to operate over 10 fps.

## 5.6 Experiments with V2V radios

**Methodology.** In this section, we replay the transmission schedules from one configuration of each scenario over a test bed with three DSRC radios; each transmission is carried over a random transceiver pair. We programmed the DSRC radios to use LTE-Direct TDMA Mode 4 [39], a listen-before-transmission mode to follow the schedule. To coordinate the radios, we designed simple handshake messages and timeout mechanisms to maintain synchronization among all radios. The precise synchronization mechanism that incurs minimum overhead is an open topic beyond the scope of this paper.

For each scenario, we record the point cloud data to transmit and the computed transmission schedule. In the simulator, the schedule is based on a theoretical model of the channel with a fixed data rate. The goal of this evaluation is to validate whether the DSRC radios can finish the scheduled transmission in time and evaluate the significance of packet loss and its impact on the transmissions.

**Results.** Figure 13 shows the end-to-end latency for each 100 ms decision interval. We played back traces with different traffic densities to see if DSRC radios can always fulfill the schedule. At low density, the number of objects to transmit is small<sup>9</sup>. The bandwidth saturates at 20 vehicles and objects get prioritized: those over 100 ms is compensated next interval. The results show that DSRC can always finish the schedule in time. Prioritization in the schedule is the reason AUTOCAST can maintain high visibility on collider and safety-critical objects. Figure 14 takes the overtaking scenario as an example to show the transmissions of each object in detail without background traffic. The scheduled transmissions are in the upper subplot and the actual DSRC transmissions on the bottom. Each shared object is represented by a line with a unique color compared to other objects in the same decision interval. Each colored line starts from the time of the beginning of the transmission, ends when the transmission completes. The x-axis represents the time in ms, the y-axis represents the time within each decision interval (100 ms).

<sup>9</sup>To evaluate DSRC under different bandwidth saturation level, we reduced the number of passive traffic participants in this setting.

If an object is lost due to channel variability and dropped packets, the solid line becomes a dashed line. In the overtaking scenario, the collider's point cloud as observed by the truck is in yellow, blue represents the truck's point cloud as observed by the collider. As the collider moves closer to the truck, both observations are larger and hence the length of yellow and blue transmission is longer. Red is the ego's point cloud as observed by the truck, and purple is the truck's point cloud as observed by the ego. Blue and purple transmissions are also scheduled: the ego and collider broadcasting the truck's (occluder's) point cloud. In theory, they need not be transmitted because the truck is visible and not relevant; AUTOCAST does transmit objects with less relevance when possible.

## 5.7 Scheduling Algorithms

We evaluate the optimality of different scheduling algorithm (*Optimal*, *FPTAS*, *Greedy*, *Agnostic*) in terms of total rewards and reward ratio, algorithm complexity and scalability with respect to the number of vehicles, and transmission delay for objects with different rewards. We conduct this set of by setting the vehicle on autopilot mode, entering and exiting an intersection from all directions.

**Optimality, Complexity, and Scalability.** We first evaluate the optimality. Figure 15 plots the total scheduled rewards when the number of vehicles varies from 5 to 40. Specifically, Greedy has less than 2% difference from Optimal while *Agnostic* has upto 40% difference from Greedy. Figure 16 further shows the computation time of the schedule with different number of vehicles. Although the total rewards are close, greedy is two orders of magnitude faster than FPTAS whereas the running time of optimal (dynamic programming) can quickly become prohibitive. Vehicular environments can be highly dynamic which requires the schedule to be computed frequently; only the proposed greedy algorithm can finish within the 100ms decision interval with 40 vehicles.

**Scheduled Delay.** The scheduled delay is measured by calculating the duration from the start of each decision interval (every 100 ms) to the time that a particular object is received. Figure 17 shows the scheduled delay of each object in a 20 cars scenario. It gives more details behind the scene, explains the reward ratio difference by showing the transmission priority. AUTOCAST's optimization always put the object with the highest normalized rewards top of the schedule which results in lower scheduled delay, whereas the latency of objects scheduled in *Agnostic* is random.

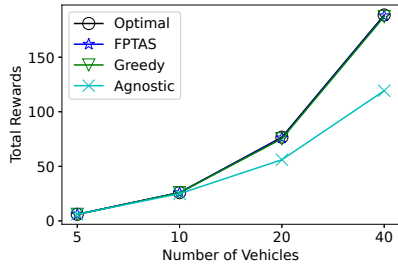


FIGURE 15: Optimality: Total Rewards

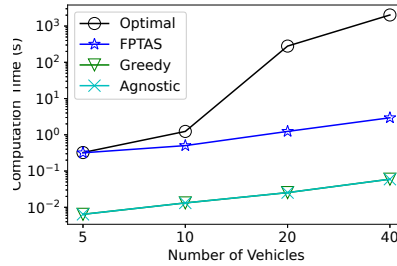


FIGURE 16: Computation Time

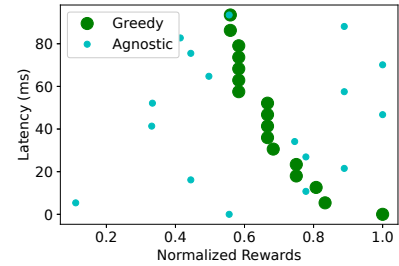


FIGURE 17: Scheduled Delay (20 Vehicles)

## 6 Related Work

**Connected Vehicles and Infrastructure:** Connected vehicles can improve the safety and reliability of self-driving cars. Communication technologies, *e.g.*, DSRC [47] and LTE-Direct [39, 67], provide capabilities to exchange information among cars by different types of transmission, *i.e.*, multicasting, broadcasting, and unicasting. Automakers are deploying V2V/V2X communications in their upcoming models [16, 17]). The academic community has started to build city-scale advanced wireless research platforms (COSMOS [7]), as well as large connected vehicle testbed in the U.S. (MCity [18]) and Europe (DRIVE C2X [77]), which gives an opportunity to explore the application feasibility of connected vehicles via V2V communications in practice.

**Sensor/Visual Information:** Processing visual information from sensors (*e.g.*, LiDAR, stereo cameras, *etc.*) is important for autonomous driving systems [56], which rely on visual information to make a on-road decisions for detection [28], tracking [57], and motion forecasting [58]. In addition, there is a large body of work on vehicle context and behavior sensing [40, 66, 71] to enhance vehicular situational awareness. AUTOCAS, and several related works discussed below, take the next step of designing how to share this information among nearby vehicles.

**Vehicle Sensor Sharing:** Prior research has attempted context sharing among vehicles. Rybicki et al. [72] discuss challenges of VANET-based approaches, and propose to leverage infrastructure to build distributed, cooperative Traffic Information Systems. Other work has explored robust inter-vehicle communication [30, 50, 70], an automotive ontology for intelligent vehicles [37], principles of community sensing that offer mechanisms for sharing data from privately held sensors [48], and sensing architectures [54] for semantic services. Motivated by “See-through” [41], several prior work on streaming video for enhanced visibility have been proposed [25, 38, 46, 53, 69].

More recent work [27, 65, 82] demonstrates the feasibility of sharing point clouds, but with limited scale and infrastructure support [22]. In AUTOCAS, we focus on enabling clusters of vehicles to share sensor information at scale in the absence of edge servers.

## 7 Limitations and Future Work

To focus on evaluating the scalability of cooperative perception and the impact on driving behavior, AUTOCAS assumes perfect pose estimation in simulation. In practice, high-definition maps result in positioning errors of 10-30 cm. We have examined the impact of this error on our end-to-end results by perturbing the vehicle location with a Gaussian noise. Because we set a 2 m threshold for near-miss, this positioning error does not affect the end-to-end results. To

demonstrate the impact on driving, we have designed three scenarios where occlusions can trigger a potential crash between vehicles. We leave small objects such as pedestrians and cyclists to future work. For vehicular connectivity, we have used DSRC radios. Our experiments demonstrate the feasibility of AUTOCAS, but DSRC technology is being phased out [8, 36], and is being replaced by other alternatives, such as C-V2X. Nevertheless, because our design can accommodate LTE-Direct TDMA Mode 4 [39], AUTOCAS is fully compatible with C-V2X radios, and should perform similarly in C-V2X equipped vehicles. Finally, V2V communication security is crucial for cooperative perception, but is currently an open problem that future work should investigate. We expect that robustness to attacks can be achieved, in part, by cross-validating contributions from different vehicles.

## 8 Conclusion

In this paper, we have designed and implemented AUTOCAS, a system that scales cooperative perception to dense traffic settings without infrastructure dependency. AUTOCAS allows vehicles to share point clouds of dynamic objects with each other, but because these can congest the wireless channel, it carefully determines which objects to transmit based on visibility and relevance. These properties are input to a distributed scheduling algorithm that determines a transmission schedule at every decision interval. For several challenging traffic scenarios, AUTOCAS significantly outperforms baseline approaches that do not employ cooperative perception, or do not prioritize objects transmissions, or rely on edge relay services. Its perception and planning pipelines have been optimized to process LiDAR data at frame rate in under 100 ms. Future work can improve perception and planning modules, experiment over LTE-V radios, design corresponding representation fusion for end-to-end control, and enhance vehicle communication security.

**Acknowledgement:** We thank our shepherd, Kasthuri Jayarajah, and the Mobisys reviewers for their valuable comments and help revising the paper. iSmartWays provided invaluable support for their DSRC and C-V2X radios.



## References

- [1] 5G Automotive Association. <https://5gaa.org/>.
- [2] *LTE-Advanced Is the Real 4G*, . <http://spectrum.ieee.org/telecom/standards/lte-advanced-is-the-real-4g>.
- [3] *The MobileBroadband LTE-Advanced Standard*, . <http://www.3gpp.org/technologies/keywords-acronyms/97-lte-advanced>.
- [4] *The MobileBroadband LTE Standard*, . <http://www.3gpp.org/technologies/keywords-acronyms/98-lte>.
- [5] *LTE-direct Trial Whitepaper*, . <https://www.qualcomm.com/media/documents/files/lte-direct-trial-whitepaper.pdf>.
- [6] Carla autonomous driving challenge. URL <https://carlachallenge.org/>.
- [7] Cloud enhanced open software defined mobile wireless testbed for city-scale deployment (cosmos). URL <https://cosmos-lab.org/>.
- [8] Modernizing the 5.9 ghz band. URL <https://docs.fcc.gov/public/attachments/DOC-367827A1.pdf>.
- [9] ismartways performance measurement, . URL <https://fccid.io/2AQQ3IM2RSE/Test-Report/FCC-Part22-4039626>.
- [10] ismartways technology, . URL <http://www.ismartways.com/>.
- [11] Numba: opensource jit compiler. URL <https://numba.pydata.org/>.
- [12] Velodyne 64-beam lidar. URL <https://velodynelidar.com/hdl-64e.html>.
- [13] Ieee standard for information technology– local and metropolitan area networks–specific requirements– part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments. *IEEE Std 802.11p-2010 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11n-2009, and IEEE Std 802.11w-2009)*, pages 1–51, July 2010.
- [14] Tr 36.785, 2016, vehicle to vehicle (v2v) services based on lte sidelink; user equipment (ue) radio transmission and reception. Technical report, 3GPP, 2016.
- [15] Tr 36.885, 2016, study on lte-based v2x services. Technical report, 3GPP, 2016.
- [16] 18 awesome innovations in the new mercedes e-class, 2016. URL <https://www.businessinsider.com/mercedes-e-class-2017-features-2016-6>.
- [17] Toyota's v2v move shows industry still interested in cars talking to each other, 2018. URL <https://www.consumerreports.org/automotive-technology/toyota-v2v-vehicle-to-vehicle-communications/>.
- [18] Ann arbor connected vehicle test environment, 2019. URL <http://www.aacvte.org>.
- [19] F. Ahmad, H. Qiu, F. Bai, and R. Govindan. CarMap: Crowdsourced Feature-Maps for Automobiles. under submission, 2019.
- [20] M. Alinci, E. Spaho, A. Lala, and V. Kolici. Clustering algorithms in manets: A review. In *2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 330–335, 2015. doi: 10.1109/CISIS.2015.47.
- [21] A. Amis, R. Prakash, T. Vuong, and D. Huynh. Max-min d-cluster formation in wireless ad hoc networks. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 1, pages 32–41 vol.1, 2000. doi: 10.1109/INFCOM.2000.832171.
- [22] J. Arulraj, A. Chatterjee, A. Daglis, A. Dhekne, and U. Ramachandran. ecloud: A vision for the evolution of the edge-cloud continuum. *Computer*, 54(5):24–33, 2021.
- [23] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall. Lte: The evolution of mobile broadband. *Comm. Mag.*, 47(4), Apr. 2009.
- [24] D. P. Bertsekas. *Dynamic Programming and Optimal Control - Vol I*. Athena Scientific, 2005.
- [25] M. Boban, W. Viriyasitavat, and O. Tonguz. Modeling vehicle-to-vehicle line of sight channels and its impact on application-layer performance. In *ACM VANET 2013*, June 2013.
- [26] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020.
- [27] Q. Chen, X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu. F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, pages 88–100, 2019.
- [28] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia. Multi-view 3d object detection network for autonomous driving. In *IEEE CVPR 2017*. IEEE, June 2017.
- [29] C. Choy, J. Gwak, and S. Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019.
- [30] T. Das, L. Chen, R. Kundu, A. Bakshi, P. Sinha, K. Srinivasan, G. Bansal, and T. Shimizu. Corecast: Collision resilient broadcasting in vehicular networks. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '18*, pages 217–229, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5720-3. doi: 10.1145/3210240.3210341. URL <http://doi.acm.org/10.1145/3210240.3210341>.
- [31] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.
- [32] D.-Z. Du, K.-I. Ko, and X. Hu. *Design and Analysis of Approximation Algorithms*. Springer, 2012.
- [33] A. Elhafi, B. Ivanovic, L. Janson, and M. Pavone. Map-predictive motion planning in unknown environments. *arXiv preprint arXiv:1910.08184*, 2019.
- [34] I. Er and W. Seah. Mobility-based d-hop clustering algorithm for mobile ad hoc networks. In *2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*, volume 4, pages 2359–2364 Vol.4, 2004. doi: 10.1109/WCNC.2004.1311457.
- [35] ETSI. Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service. Technical Report REN/TTS-0010019, 2014.
- [36] FCC. Use of the 5.850-5.925 ghz band. URL <https://www.federalregister.gov/documents/2021/05/03/2021-08802/use-of-the-5850-5925-ghz-band>.
- [37] M. Feld and C. Müller. The automotive ontology: Managing knowledge inside the vehicle and sharing it between cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '11*, pages 79–86, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1231-8. doi: 10.1145/2381416.2381429. URL <http://doi.acm.org/10.1145/2381416.2381429>.
- [38] M. Ferreira, P. Gomes, M. K. Silveria, and F. Vieira. Augmented reality driving supported by vehicular ad hoc networking. In *IEEE ISMAR 2013*, October 2013.
- [39] L. Gallo and J. Harri. Short paper: A lte-direct broadcast mechanism for periodic vehicular safety communications. In *IEEE Vehicular Networking Conference 2013*, pages 166–169. IEEE, Dec 2013. doi: 10.1109/VNC.2013.6737604.
- [40] T. Gandhi and M. M. Trivedi. Motion based vehicle surround analysis using an omni-directional camera. In *Intelligent Vehicles Symposium, 2004 IEEE*, pages 560–565. IEEE, 2004.
- [41] P. Gomes, F. Vieira, and M. Ferreira. The see-through system: From implementation to test-drive. In *2012 IEEE Vehicular Networking Conference (VNC)*, pages 40–47. IEEE, Nov 2012. doi: 10.1109/VNC.2012.6407443.
- [42] O. Grembek, A. A. Kurzhanskiy, A. Medury, P. Varaiya, and M. Yu. An intelligent intersection. *CoRR*, abs/1803.00471, 2018. URL <http://arxiv.org/abs/1803.00471>.
- [43] J. Guanetti, Y. Kim, and F. Borrelli. Control of connected and automated vehicles: State of the art and future challenges. *Annual Reviews in Control*, 45:18–40, 2018. ISSN 1367-5788. doi: 10.1016/j.arcontrol.2018.04.011. URL <http://dx.doi.org/10.1016/j.arcontrol.2018.04.011>.
- [44] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.
- [45] L. Janson and M. Pavone. Fast marching trees: a fast marching sampling-based method for optimal motion planning in many dimensions - extended version. *CoRR*, abs/1306.3532, 2013. URL <http://arxiv.org/abs/1306.3532>.
- [46] J. Judvaitis, A. Hermanis, K. Nesenbergs, R. Cacus, I. Homjakovs, and K. Sudars. Object transparent vision combining multiple images from different views. *Automatic Control and Computer Sciences*, 49(5):313–320, Sep 2015. ISSN 1558-108X. doi: 10.3103/S0146411615050053. URL <https://doi.org/10.3103/S0146411615050053>.
- [47] J. B. Kenney. Dedicated short-range communications (dsrc) standards in the united states. *Proceedings of the IEEE*, 99(7):1162–1182, July 2011. ISSN 0018-9219. doi: 10.1109/JPROC.2011.2132790.
- [48] A. Krause, E. Horvitz, A. Kansal, and F. Zhao. Toward community sensing. In *Proceedings of the 7th International Conference on Information Processing in Sensor Networks, IPSN '08*, pages 481–492, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-0-7695-3157-1. doi: 10.1109/IPSN.2008.37. URL <http://dx.doi.org/10.1109/IPSN.2008.37>.
- [49] T. Krisher and J. Lowy. Tesla driver killed in crash while using car's 'autopilot'. *ASSOCIATED PRESS (June 30, 2016)*.
- [50] S. Kumar, L. Shi, N. Ahmed, S. Gil, D. Katabi, and D. Rus. Carspeak: A content-centric network for autonomous driving. In *ACM SIGCOMM 2012*, August 2012.
- [51] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom. Pointpillars: Fast encoders for object detection from point clouds, 2019.
- [52] S. LaValle. Rapidly-exploring random trees : a new tool for path planning. 1998.
- [53] P. Lindemann and G. Rigoll. Examining the impact of see-through cockpits on driving performance in a mixed reality prototype. In *ACM AutomotiveUI 2017*, September 2017.
- [54] J. Liu and F. Zhao. Towards semantic services for sensor-rich information systems. In *2nd International Conference on Broadband Networks, 2005.*, pages 967–974 Vol. 2, Oct 2005. doi: 10.1109/ICBN.2005.1589709.
- [55] W. Luo, B. Yang, and R. Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [56] W. Luo, B. Yang, and R. Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *IEEE CVPR 2018*. IEEE, June 2018.
- [57] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang. Hierarchical convolutional features for visual tracking. In *IEEE ICCV 2015*. IEEE, December 2015.
- [58] W.-C. Ma, D.-A. Huang, N. Lee, and K. M. Kitani. Forecasting interactive dynamics of pedestrians with fictitious play. In *IEEE CVPR 2017*. IEEE, June

- 2017.
- [59] G. Mattyus, S. Wang, S. Fidler, and R. Urtasun. Hd maps: Fine-grained road segmentation by parsing ground and aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
  - [60] W. G. Najm, R. Ranganathan, G. Srinivasan, J. D. Smith, S. Toma, E. Swanson, A. Burgett, et al. Description of light-vehicle pre-crash scenarios for safety applications based on vehicle-to-vehicle communications. Technical report, United States. National Highway Traffic Safety Administration, 2013.
  - [61] National Highway Traffic Safety Administration. Fatality Analysis Reporting System. <https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf>. Talk at SXSW Interactive 2016.
  - [62] M. Ni, Z. Zhong, and D. Zhao. Mphc: A mobility prediction-based clustering scheme for ad hoc networks. *IEEE Transactions on Vehicular Technology*, 60(9): 4549–4559, 2011. doi: 10.1109/TVT.2011.2172473.
  - [63] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics)*. Wiley-Interscience, New York, NY, USA, 2007. ISBN 0470171553.
  - [64] H. Qiu, F. Ahmad, F. Bai, M. Gruteser, and R. Govindan. Avr: Augmented vehicular reality. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, MobiSys '18, pages 81–95, Munich, Germany, 2018. ACM.
  - [65] H. Qiu, F. Ahmad, F. Bai, M. Gruteser, and R. Govindan. Avr: Augmented vehicular reality. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, pages 81–95, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5720-3.
  - [66] H. Qiu, J. Chen, S. Jain, Y. Jiang, M. McCartney, G. Kar, F. Bai, D. K. Grimm, M. Gruteser, and R. Govindan. Towards robust vehicular context sensing. *IEEE Transactions on Vehicular Technology*, 67(3):1909–1922, 2018.
  - [67] Qualcomm. Lte direct proximity services, 2019. URL <https://www.qualcomm.com/invention/technologies/lte/direct>.
  - [68] A. H. Qureshi, A. Simeonov, M. J. Bency, and M. C. Yip. Motion planning networks. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 2118–2124. IEEE, 2019.
  - [69] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park, and I. S. Kweon. A real-time augmented reality system to see-through cars. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2395–2404, Nov 2016. ISSN 1077-2626. doi: 10.1109/TVCG.2016.2593768.
  - [70] D. Reichardt, M. Miglietta, L. Moretti, P. Morsink, and W. Schulz. Cartalk 2000: Safe and comfortable driving based upon inter-vehicle-communication. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, pages 545–550. IEEE, 2002.
  - [71] A. Ruta, F. Porikli, S. Watanabe, and Y. Li. In-vehicle camera traffic sign detection and recognition. *Machine Vision and Applications*, 22(2):359–375, 2011.
  - [72] J. Rybicki, B. Scheuermann, W. Kiess, C. Lochert, P. Fallahi, and M. Mauve. Challenge: Peers on wheels - a road to new traffic information systems. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, MobiCom '07, pages 215–221, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-681-3. doi: 10.1145/1287853.1287879. URL <http://doi.acm.org/10.1145/1287853.1287879>.
  - [73] M. Saxena, N. Phate, K. Mathai, and M. Rizvi. Clustering based energy efficient algorithm using max-heap tree for manet. In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pages 123–127, 2014. doi: 10.1109/CSNT.2014.33.
  - [74] S. Shi, X. Wang, and H. Li. Pointcrnn: 3d object proposal generation and detection from point cloud. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
  - [75] M. Siam, S. Elkerdawy, M. Jagersand, and S. Yogamani. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2017. doi: 10.1109/itsc.2017.8317714. URL <http://dx.doi.org/10.1109/ITSC.2017.8317714>.
  - [76] A. Sinha and E. Modiano. Throughput-optimal broadcast in wireless networks with point-to-multipoint transmissions. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '17, pages 3:1–3:10, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4912-3.
  - [77] R. Stahlmann, A. Festag, A. Tomatis, I. Radusch, and F. Fischer. Starting european field tests for car-2-x communication: the drive c2x framework. In *ITS World Congress and Exhibition*, Oct 2011.
  - [78] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *European Conference on Computer Vision*, pages 605–621. Springer, 2020.
  - [79] T. Yin, X. Zhou, and P. Krähenbühl. Center-based 3d object detection and tracking, 2021.
  - [80] J. Zhang and S. Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems*, volume 2, page 9, 2014.
  - [81] J. Zhang and S. Singh. Visual-lidar odometry and mapping: Low-drift, robust, and fast. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2174–2181. IEEE, 2015.
  - [82] X. Zhang, A. Zhang, J. Sun, X. Zhu, Y. E. Guo, F. Qian, and Z. M. Mao. Emp: edge-assisted multi-vehicle perception. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 545–558, 2021.
  - [83] Y. Zhou and O. Tuzel. Voxnet: End-to-end learning for point cloud based 3d object detection, 2017.