拥堵匝道汇入场景中用于安全高效网联自动驾驶汽车变道的协同安全屏障

Bharathkumar Hegde[1],M´elanie Bouroche[1]

arXiv:2602.10007v1 [cs.RO] 10 Feb 2026

摘要——密集交通中的变道对网联与自动驾驶车辆（CAV）而言是一项重大挑战。现有换道控制器主要侧重于确保安全或协同提升交通效率，但尚未同时考虑这两个相互冲突的目标。为此，我们提出多智能体安全屏蔽器（MASS），其采用控制屏障函数（CBF）设计，以实现安全、协同的变道。MASS通过简单算法构建图形式的交互拓扑，捕捉CAV之间的多智能体交互，从而实现协作。此外，我们在先进的多智能体强化学习（MARL）换道控制器中集成MASS以确保安全，并定义定制奖励函数以优先提升效率。据此，我们提出一种名为MARL-MASS的换道控制器，并在拥堵匝道汇入仿真中对其进行评估。结果表明，MASS能够在严格遵守安全约束的同时，实现具有安全保证的协同换道。此外，所提出的定制奖励函数提升了采用安全屏障训练的MARL策略的稳定性。总体而言，MARL-MASS通过鼓励探索协同换道策略，同时遵守安全约束，有效平衡了拥堵交通中确保安全与提升交通效率之间的权衡。

I. 引言

缓解交通拥堵是交通流管理面临的主要挑战之一。交通拥堵会造成广泛影响，包括行程时间可靠性降低、燃油消耗增加、供应链成本上升、交通流效率低下以及事故率上升 [1]。例如，《2023年城市交通报告》估计，美国交通拥堵造成的总成本约为 2240 亿美元 [2],；这一估算基于每年 85 亿小时的交通延误和 33 亿加仑的燃油浪费。在拥堵合流场景中，高效变道有助于改善交通流，从而预防或缓解拥堵。然而，在密集交通中尝试变道颇具挑战，因为这是一项安全关键任务，涉及多辆网联自动驾驶车辆以高速独立且同时作出决策。

不当变道可能导致碰撞，从而造成重大损害和人员伤亡。重要的是，车辆的安全不仅取决于其控制决策，还取决于周围车辆可能引发碰撞的自私决策。例如，与轻度和中度交通相比，在密集交通中更难找到汇入机会，如图1所示。在此类场景中，网联自动驾驶车辆（网联自动驾驶汽车）可以执行协同变道，以提高交通效率。因此，

[1]都柏林圣三一学院计算机科学与统计学院，爱尔兰hegdeb@tcd.ie, Melanie.Bouroche@tcd.ie

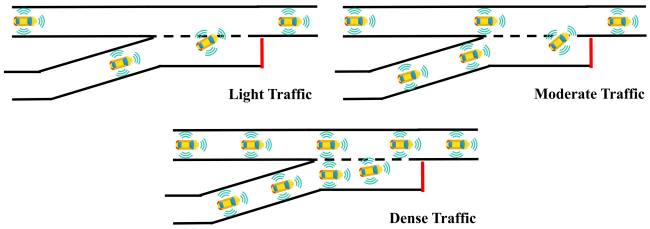


图 1. 高速公路合流路段不同交通流量等级示意图

网联自动驾驶车辆技术已引起研究界的关注，因为它们使车辆能够借助车联万物（V2X）通信做出协同控制决策 [3]。

为了实现网联自动驾驶车辆之间的协作，一种多智能体形式的强化学习（RL），即多智能体强化学习（MARL），正因其能够为多个在高度动态的环境中同时且独立行动的智能体学习控制策略的潜力，而在研究领域日益受到关注 [4], [5]。多种形式的 MARL 已被用于设计换道控制器，包括深度Q网络（DQNs） [6], [7], [8], 和行动者-评论家网络（ACN） [9], [10], [11]。其中，采用集中式安全屏障的 MARL 换道控制器称为 MARL-CS，它使用近端策略优化的多智能体扩展版本（MAPPO），通过 ACN 训练策略。MARL-CS 因能够学习稳定且可复现的策略，从而在高速公路合流场景中高效地做出变道决策而闻名 [11]。此外，该换道控制器可通过开源代码库获取。因此，MARL-CS 被视为面向网联自动驾驶汽车的最先进 MARL 换道控制器。尽管这些 MARL 换道控制器的首要目标是提高效率并尽量减少碰撞，但它们并不提供安全保证。

控制屏障函数（CBF）已被认定为确保网联自动驾驶车辆中分布式多智能体强化学习策略安全性的合适方案 [12]。事实上，由于多智能体强化学习策略可能作出不安全的控制决策，因此可以利用控制屏障函数定义的安全约束覆盖这些不安全动作，确保智能体保持在安全状态 [13], [14]。为确保网联自动驾驶车辆多智能体强化学习换道控制器的安全性，研究人员提出了混合安全屏障（HSS），将优化方法与基于规则的方法相结合，以有效约束加速度和转向角输入 [15]。尽管车辆会独立且同时变道，HSS 仍能确保车辆安全；它采用谨慎策略，假设其他车辆会作出最坏情况控制决策。集成了 HSS 的多智能体强化学习换道控制器，即 MARL-HSS，在轻度至中度交通状况下提升了交通效率，同时确保安全。然而，随着交通变得更加拥堵，MARL-HSS

由于安全变道的机会有限，难以收敛到稳定的策略，导致交通效率低下。

相反，网联自动驾驶汽车可以通过协同控制决策，即使在密集交通中也能提升交通效率。将网联自动驾驶汽车视为多智能体系统（MAS），通过基于时隙的方法、 [16], 博弈论 [17],和基于优先级的方法 [11], [18]等协作方式，可以高效地做出变道决策。然而，这些方法依赖集中式服务器来协调匝道汇入过程中车辆之间的协作，因此限制了其适用性和可扩展性。作为替代方案，分布式协作方法通常通过图来描述多智能体动态，这类图称为交互拓扑 [19]–[21]。这种方法在设计网联自动驾驶汽车安全屏障方面的潜在应用仍有待研究。

本文探讨如何将交互拓扑应用于构建协同安全屏障，并将其与多智能体强化学习变道控制器相结合，以缓解匝道汇入处的交通拥堵，同时确保安全。本文的主要贡献如下：

- 我们提出了一种新型协同安全屏障，称为多智能体安全屏蔽器（MASS）。该屏蔽器基于 HSS，通过考虑其他网联自动驾驶汽车的控制决策，利用控制障碍函数约束评估安全控制输入。我们将 MASS 与多智能体强化学习相结合，构建一种面向网联自动驾驶汽车的协同控制器，称为 MARL-MASS，可通过 https://github.com/hkbharath/ MARL-MASS 获取。

- 为促进 MASS 中的协作，我们提出了一种算法，用于构建高速公路合流场景中网联自动驾驶车辆的交互拓扑。该拓扑表示为一个图，用于评估联合安全控制动作。

- 我们为集成于 MASS 的多智能体强化学习控制器明确设计了定制奖励函数，以优先实现交通效率优化，同时由安全屏障确保安全。

- 我们在类 Gym 仿真平台上对 MARL-MASS 进行了实证评估，结果表明，与基线多智能体强化学习控制器相比，它在缓解密集交通的同时能够遵守安全约束，表现更优。

II. 背景

本节介绍车辆控制、多智能体强化学习和控制屏障函数的背景知识，为所提出的方法奠定基础。

A. 车辆控制

通常，车辆控制具有多个对应于不同驾驶任务的层级 [22]。其中，行为层做出离散决策，以变道或沿当前车道行驶。随后，运动规划层确定控制参考，以执行所需的驾驶机动，从而落实每一项行为决策。这些驾驶机动随后由车辆

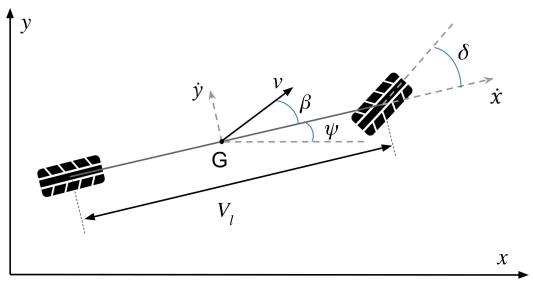


图 2. 运动学自行车模型

控制器基于定义自动驾驶汽车（自动驾驶车辆）动力学的车辆模型。

自动驾驶汽车动力学可能受到多种因素的影响，从车辆控制到天气状况皆是如此，因此难以建模。为简化起见，通常使用运动学自行车模型来模拟自动驾驶车辆运动 [23]，，我们也使用该模型演示所提出的安全屏障。运动学自行车模型将两个前轮视为一个车轮，后轮也同样处理，如图 2 所示。以下方程定义了车辆模型，用于更新车辆的状态变量：

$$\begin{aligned} \dot{x} &= v_x, \quad \dot{y} = v_y, \quad \dot{v}_x = a\cos(\psi+\beta), \\ \dot{v}_y &= a\sin(\psi+\beta), \quad \dot{\psi} = \frac{2\ v}{V_l}\sin\beta, \end{aligned} \tag{1}$$

其中， β 是根据转向角输入 δ计算得出的质心处侧偏角：

$$\beta = \tan^{-1}\left(\frac{1}{2}\tan\delta\right) \tag{2}$$

B. 用于变道的多智能体强化学习

在一种用于匝道汇入的最先进的多智能体强化学习变道控制器 MARL-CS 中，变道决策采用离散行为决策的形式，例如 {向右、向左、保持车道、加速、减速}，并依据观测到的状态空间作出；该状态空间由以下状态变量组成x ： 车辆的纵向位置、 y 车辆的横向位置、 v_x 车辆的纵向速度、 v_y 车辆的横向速度，以及 ψ 车辆相对于道路的航向。每个离散时间步 t，网联自动驾驶汽车通过结合各自的决策作出联合行为决策。

MARL-CS 中使用的奖励函数会在智能体实现个人目标和协作目标时给予奖励。网联自动驾驶车辆 i 的奖励定义如下：

$$r_i = w_{\mathrm{c}} r_{\mathrm{c}} + w_{\mathrm{s}} r_{\mathrm{s}} + w_{\mathrm{h}} r_{\mathrm{h}} + w_{\mathrm{m}} r_{\mathrm{m}} \tag{3}$$

其中，网联自动驾驶车辆因避免碰撞而获得 r_{c}的奖励、因保持期望速度而获得 r_{s}的奖励、因保持期望车头时距而获得 r_{h}的奖励，以及因最小化汇入等待时间而获得r_{m}的奖励。每项奖励都乘以权重 w_* ，该权重可调，以优先考虑不同目标。为奖励协作目标，来自 n个观测

感知范围内的车辆被合并，从而得到如下总体奖励函数：

$$R_i = \frac{1}{n}\sum_{j=0}^{n} r_j \tag{4}$$

C. 用于运动规划的控制障碍函数

控制障碍函数通常用于约束非线性控制系统，例如机器人控制器或网联自动驾驶车辆，以确保安全控制决策 [24], [25]。考虑一个离散时间非线性控制系统，该系统基于非驱动动力学 f： $\mathcal{S} \to \mathcal{S}$和受控动力学$g$： $\mathcal{S} \to \mathbb{R}^{n,\ m}$定义下一状态 s_{t+1}，其中 n 和 m 分别为状态空间 $\mathcal{S}$ 和动作空间 U中的变量数量。该状态转移动力学可定义如下，

$$s_{t+1} = f(s_t) + g(s_t)u_t \tag{5}$$

其中， $s_t \in \mathcal{S}$ 为系统状态， $u_t \in U$ 为时间步 t时的控制动作。 f 和 g 根据已知系统动力学定义，并且它们是局部利普希茨连续的，换言之，它们是受最大变化速率限制的连续函数。

利用连续可微的控制障碍函数 h： $\mathcal{S} \to \mathbb{R}$，可以定义安全集 C： $\{s_t \in \mathcal{S}:\ h(s_t) \geq 0\}$，其中 $h(s_t)$是一个形式为 $h(s_t) = p^{\mathrm{T}}s_t + q$ 的仿射屏障函数，用于定义非线性控制系统的安全约束。利用该约束，可以构建一个优化问题，以评估确保方程（5）中控制系统安全的最小控制动作 u_t [13]。该优化问题的详细推导见 [13] 和 [24]。

在 MARL-HSS [15] 中，展示了将 CBF 应用于保障网联自动驾驶车辆的 MARL 换道控制器的安全性。在此方法中，HSS 会覆盖可能不安全的低层动作，例如 v_{e}；这些动作是为执行多智能体强化学习策略针对自车作出的高层行为决策而评估的。虽然 HSS 会考虑观测车辆的状态，但在评估满足 CBF 约束 $h(s_t)$ 所需的最小修正值 v^{cbf} 时，假设这些车辆会作出最坏情况控制决策 $v_{\mathrm{o}}^{\mathrm{wc}}$。这样，无论周围车辆的控制决策如何，都能确保满足安全约束。因此，安全动作 v^{safe} 定义为

$$v^{\mathrm{safe}} = \underbrace{\begin{bmatrix} v_{\mathrm{e}} & v_{\mathrm{o}}^{\mathrm{wc}} \end{bmatrix}}_{v} + \underbrace{\begin{bmatrix} v_{\mathrm{e}}^{\mathrm{cbf}} & 0 \end{bmatrix}}_{v^{\mathrm{cbf}}} \tag{6}$$

在 HSS 中，纵向安全约束基于相对于观测到的前车的车头时距阈值 τ，使用控制屏障函数 $h_{\mathrm{ol}}(s_t)$ 定义如下

$$h_{\mathrm{ol}}(x_t) = \Delta x_{\mathrm{ol}} - x^{\mathrm{safe}} \tag{7}$$

其中 $x^{\mathrm{safe}} = \tau * v_{\mathrm{e}}$ 和 Δx_{ol} 是自车与观测到的前车之间的纵向距离。类似地，HSS采用相对于相邻车道前车和后车的控制屏障函数约束 $h_{\mathrm{oal}}(s_t)$和 $h_{\mathrm{oar}}(s_t)$作为横向安全约束。仅当横向约束得到满足时，HSS才允许变道，以确保安全的横向控制决策 [15]。

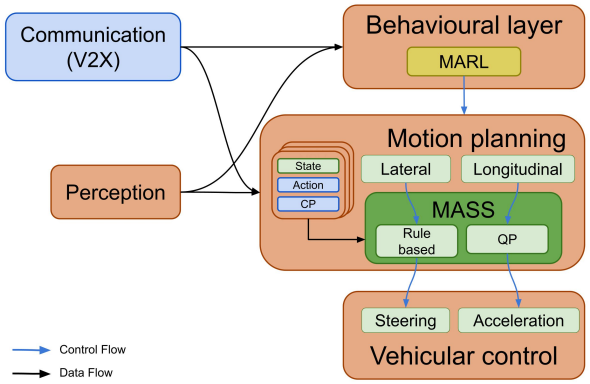


图 3. MARL-MASS 车辆控制器架构

III. 所提方法

本节阐述本文的主要贡献，由此形成MARL-MASS，一种将多智能体强化学习行为层与MASS相集成的安全高效换道控制器，如图3所示。第一项贡献MASS在第III-A节中设计。为识别网联自动驾驶汽车多智能体系统中的交互依赖关系，第III-B节提出一种用于构建交互拓扑的算法。由于协同安全屏障确保了安全性，多智能体强化学习策略可以完全专注于效率提升。因此，第III-C节为带有安全屏障的多智能体强化学习换道控制器提出了合适的奖励函数。

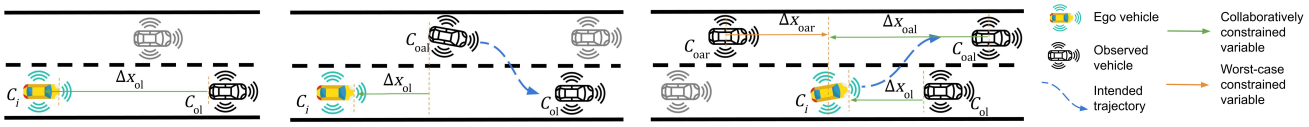
A. 多智能体安全屏蔽器

MASS 是一种安全屏障，使网联自动驾驶车辆能够在每个时间步，通过在运动规划层采取安全控制动作，执行多智能体强化学习策略在行为层提出的行为动作。为构建 MASS，定义给定时间步下车辆纵向运动的控制系统 (5) 如下

$$\begin{bmatrix} x_{\mathrm{e}} \\ x_{\mathrm{o}} \end{bmatrix} = \underbrace{\begin{bmatrix} x_{\mathrm{e}} \\ x_{\mathrm{o}} \end{bmatrix}}_{f(x)} + \underbrace{\begin{bmatrix} \cos\left(\psi_{\mathrm{e}} + \beta_{\mathrm{e}}\right) \\ g(x_o) \end{bmatrix}}_{g(x)} \underbrace{\begin{bmatrix} v_{\mathrm{e}} & v_{\mathrm{o}}^{\mathrm{safe}} \end{bmatrix}}_{u_t} \tag{8}$$

其中， ψ_{e} 和 β_{e} 分别是车辆模型 (1) 中定义的自车航向角和侧偏角。 x_{e} 和 v_{e} 分别是自车的纵向位置和速度。类似地， x_{o} 和 $v_{\mathrm{o}}^{\mathrm{safe}}$分别是观测车辆的纵向位置和速度；自车依赖该观测车辆来评估安全控制决策。 $g\left(x_{\mathrm{o}}\right)$是根据观测车辆的受控动力学评估得出的控制参数。自车依赖观测车辆共享状态 x_{o}、安全控制输入 $v_{\mathrm{o}}^{\mathrm{safe}}$和控制参数 $g\left(x_{\mathrm{o}}\right)$。虽然该控制系统针对单个观测车辆定义，但也可以进一步扩展，以考虑多个观测车辆。

由于观测车辆的状态是动态的，因此在控制系统中考虑其安全控制输入 $v_{\mathrm{o}}^{\mathrm{safe}}$。标称控制输入 u_t 可视为多智能体控制输入 v^{ma}，因为它由自车的标称控制输入 v_{e} 和



(a) 跟随车道　(b) 相邻车辆变道　(c) 变道车辆

图 4. 采用分散式控制屏障函数安全屏障的网联自动驾驶汽车作用依赖关系

观测车辆的安全控制输入 $v_{\mathrm{o}}^{\mathrm{safe}}$组成。CAV 中的 MASS 以最小干预的方式覆盖标称控制输入 v_{e}，从而评估安全控制输入 v^{safe}，如下所示:

$$v^{\mathrm{safe}} = \underbrace{\begin{bmatrix} v_{\mathrm{e}} & v_{\mathrm{o}}^{\mathrm{safe}} \end{bmatrix}}_{v^{\mathrm{ma}}} + \underbrace{\begin{bmatrix} v_{\mathrm{e}}^{\mathrm{cbf}} & 0 \end{bmatrix}}_{v^{\mathrm{cbf}}} \tag{9}$$

使用此控制输入，为多智能体网联自动驾驶车辆控制系统构建以下优化问题（式 8），以求得确保自车安全的 v^{cbf} 最小值。

$$\begin{aligned} v^{\mathrm{cbf}} &= \arg\min_{v^{\mathrm{safe}}} ||v^{\mathrm{safe}} - v^{\mathrm{ma}}||_2 + K_\epsilon \epsilon \quad \text{s.t} \quad A v^{\mathrm{safe}} \leq b, \\ \text{where } b &= [b_1, b_2, ..., b_k], \\ \text{with } b_i &= p_i^{\mathrm{T}} f(s_t) + p_i^{\mathrm{T}} (\eta - 1) s_t \ + \eta q_i + A v^{\mathrm{ma}} \end{aligned} \tag{10}$$

需要注意的是，MASS 依赖于观测车辆的状态、安全控制输入和控制参数来评估安全控制输入。这一要求使观测车辆与自车之间产生依赖关系。为处理网联自动驾驶车辆之间的此类依赖关系，我们构建了交互拓扑，以便在运动规划层评估联合安全控制输入。

B. 网联自动驾驶车辆的交互拓扑

针对穿越合流路段的网联自动驾驶车辆多智能体系统，定义交互拓扑，以确定自车在做出自身安全控制决策时所依赖的周围车辆。周围车辆是观测车辆的子集，例如前车 C_{ol}、相邻前导车辆 C_{oal}和相邻后车 C_{oar}。交互拓扑是一个图 $\mathcal{G}$，表示为一个列表，其中每辆网联自动驾驶车辆 C_i 都映射到其所依赖的一组父车辆 $\mathcal{P}$ C_i 。交互拓扑支持多智能体系统中的协同控制决策，并根据网联自动驾驶车辆的行为动作进行定义。

如果自车 C_i 正在沿车道行驶，其控制动作取决于通信范围内 C_{ol} 紧邻的前车，如图 4a 所示。基于这种依赖关系，自车可以根据前车 C_{ol}作出的控制决策，作出最优控制决策，以保持安全距离 Δx_{ol}。如果相邻车辆 C_{oal} 正尝试变道，以在前车 C_{ol}之前汇入当前车道，则认为自车依赖于该相邻车辆，如图 4b 所示。在这种情况下，自车将相邻车辆C_{oal} 视为前车，以保持安全的纵向前车间距 Δx_{ol}。这种依赖关系使相邻车辆能够协同，从而允许车辆在已经开始变道的情况下安全完成变道。请注意，安全屏障仅允许在目标车道存在安全间隙时开始变道。因此，这种依赖关系确保尝试变道的车辆所作决策优先于沿车道行驶的车辆所作决策。

如果自车 C_i 正在变道，那么它依赖于图4c所示的前车 C_{ol} 和目标车道上的相邻前导车辆 C_{oal}。根据这些车辆的状态和控制动作，自车可以做出安全控制决策，确保安全完成变道操作，同时分别与前车C_{ol} 和相邻车辆 C_{oal} 保持安全纵向距离 Δx_{ol} 和 Δx_{oal}。为避免死锁，交互拓扑中不考虑相邻车道上的后车 C_{oar}。然而，该车的状态变量会用于约束自车，防止其在横向约束下做出不安全的变道，并假设采用最坏情况控制决策。这种依赖关系使网联自动驾驶车辆能够与周围车辆协同做出控制决策，确保变道时的安全。到达目标车道后，自车可以继续沿车道行驶，同时遵循先前为沿车道行驶车辆定义的依赖关系。

算法 1 交互拓扑

输入： $\mathcal{N}$，所有车辆列表
输出： $\mathcal{G}$，交互拓扑

$\mathcal{G} \leftarrow \varnothing$
对所有 C_i 属于 $\mathcal{N}$ **执行**
　$C_{\mathrm{ol}} \leftarrow$ 相对于 C_i的前车
　$C_{\mathrm{oal}} \leftarrow$ 相对于 C_i的相邻前导车辆
　$\mathcal{P} \leftarrow \{C_{\mathrm{ol}}\}$
　如果 C_{oal} 变道至当前车道并位于 C_{ol} 后方**则**
　　$\mathcal{P} \leftarrow \{C_{\mathrm{oal}}\}$
　结束如果
　如果 C_i 正在变道**且** C_{oal} 存在于目标车道 **then**
　　$\mathcal{P} \leftarrow \mathcal{P} \cup \{C_{\mathrm{oal}}\}$
　结束如果
　$\mathcal{G} \leftarrow \mathcal{G} \cup \{ (C_i, \mathcal{P}) \}$
13: </style}循环结束
返回 $\mathcal{G}$

生成集成拓扑的步骤总结于算法 1 中。该算法以场景中的一组$\mathcal{N}$ 网联自动驾驶汽车作为输入，并返回交互拓扑 $\mathcal{G}$。该拓扑可供 MASS 使用，以确保

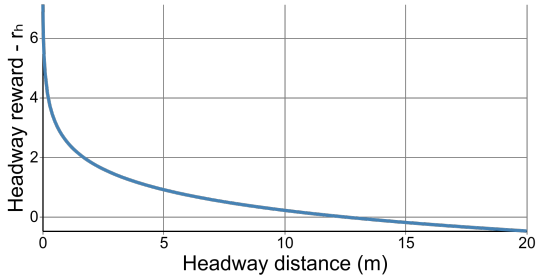


图 5. 速度为 25m/s 的车辆的车头时距奖励示意图

通过相对于周围车辆评估安全约束，分别确保网联自动驾驶汽车的纵向和横向安全。遵循交互拓扑所定义的依赖关系，可以集体确保所有网联自动驾驶汽车的安全，因为每辆网联自动驾驶汽车都会确保相对于其父车辆的安全。

C. 使用 *MASS* 进行匝道汇入的定制奖励

由于所提出的安全屏障（III-A）可确保始终保持安全车头时距，因此，定制个体奖励函数时，我们忽略碰撞惩罚，并鼓励网联自动驾驶汽车协作汇入、提高速度并缩短车头间距。为此，我们定义如下定制奖励函数，以取代当前奖励函数（3）：

$$r_i = w_{\mathrm{h}} r_{\mathrm{h}} + w_{\mathrm{s}} r_{\mathrm{s}} + w_{\mathrm{m}} r_{\mathrm{m}} \tag{11}$$

车头时距奖励 r_{h} 定义如下，用于惩罚每个智能体保持较大的车头间距：

$$r_{\mathrm{h}} = -\log \frac{\Delta x_{\mathrm{ol}}}{\tau * v_{\mathrm{e}}} \tag{12}$$

其中， Δx_{ol} 是与前车的车头间距，τ 是期望时间车头时距， v_{e} 是自车的速度。当自车更接近前车时，该奖励会增加；距离前车更远时，该奖励会减少。例如，假设一辆车以 25 m/s 的速度行驶，且期望时间车头时距为 $\tau = 0.5$ s，则车头时距奖励 r_{h} 随车头间距变化的情况如图 5 所示。该图表明，智能体保持较小的车头间距可以获得更高的奖励。

速度奖励 r_{s} 的计算方式如下

$$r_{\mathrm{s}} = \min\left\{\frac{v_{\mathrm{e}} - v_{\min}}{v_{\max} - v_{\min}}, 1\right\} \tag{13}$$

其中， $v_{\min}$ 和 $v_{\max}$ 分别为高速公路上的最低和最高期望速度。这些数值根据 MARL-CS [11] 中使用的数值，分别设定为 $v_{\min}$ =10 m/s 和 $v_{\max} = 30$ m/s。该奖励会鼓励智能体以较高速度行驶，接近 $v_{\max}$ ，从而提高交通效率。

合流奖励 r_{m} 会对在合流车道中等待的智能体施加惩罚。该奖励定义为

$$r_{\mathrm{m}} = -\exp\left\{\frac{-(x_{\mathrm{m}} - L)^2}{10L}\right\} \tag{14}$$

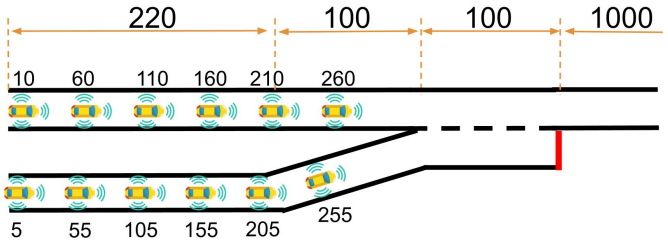


图 6. 匝道合流场景 [15]

其中， x_{m} 表示车辆在合流车道上行驶的距离， L 表示合流车道的长度。采用该奖励后，智能体在未并入高速公路 [11] 的情况下，在合流车道上行驶得越远，所受惩罚就越大。

请注意，MASS 会限制车辆与前车距离过近，但车辆可能会减速，以保持较小的车头间距并获得更高的车头时距奖励。然而，低速会受到速度奖励 r_{s}的惩罚。此外，车辆必须正确把握变道机会，尽快完成合流，以避免受到更高的合流奖励惩罚。因此，采用这一奖励函数时，智能体需要根据交通状况权衡多个目标，决定行为动作，例如提高速度、保持最小车头间距以及高效合流。

IV. 评估与结果

本节评估所提出的方法MARL-MASS，以展示安全保证、分析MARL训练稳定性，并评估训练策略。我们将换道控制器与基线MARL换道控制器进行比较，例如不安全*MARL*、*MARL-CS*以及*MARL-HSS*。不安全MARL换道策略在不使用任何安全屏障的情况下进行训练，以估计多智能体强化学习所能实现的安全与效率收益程度。MARL-CS和MARL-HSS换道控制器将多智能体强化学习与额外的安全屏障相结合，例如集中式基于优先级的安全屏障 [11],和HSS [15]。MARL-HSS已展示了安全保证，而MARL-CS则通过协同换道提高了交通效率。因此，这些换道控制器被用作基线，以评估所提出的协同安全屏障MASS。

为评估这些方法，首先定义用于训练多智能体强化学习的仿真设置（第IV-A节）。随后，通过MARL训练期间观测到的最小时间车头时距来展示MASS提供的安全保证（第IV-B节）。接下来，给出学习曲线，以分析在附加安全屏障条件下多智能体强化学习策略的训练稳定性（第IV-C节）。最后，对采用MARL-MASS、MARL-HSS和MARL-CS训练的策略进行比较，以分析安全与效率之间的权衡。

A. 评估场景与仿真设置

采用匝道合流场景训练和评估 MARL 换道控制器。这与基准 MARL-HSS 中使用的

场景相同，但交通密度更高：如图 6 所示，7–11 辆网联自动驾驶汽车通过一段 100 m 长的合流路段。此处考虑密集交通，而非 MARL-HSS 中考虑的轻度和中度交通，以展示 MARL-MASS 控制器协同创造间隙、从而在拥堵交通中实现安全换道的能力。在 MASS 中，控制障碍函数约束设置为与前车保持 0.5 s 的时间车头时距，因为与自动驾驶车辆或人工驾驶车辆相比，网联自动驾驶汽车预计会保持更小的车头时距。假设网联自动驾驶汽车在 180 m 范围内进行通信。

通过将广泛用于评估基于 AI 的自动驾驶车辆控制器的*highway-env* 模拟器 [26]，与多智能体强化学习算法（MAPPO）和 MASS 集成，对其进行了扩展，以便训练 MARL-MASS 策略。训练时，定制奖励函数的奖励权重设置为 $w_{\mathrm{h}}=1$、$w_{\mathrm{s}}=4$ 和 $w_{\mathrm{m}}=8$，以契合其优先级并提高效率。具体而言，w_{m} 加倍，以优先考虑协同合流；w_{s} 提高，以奖励更高的速度；w_{h} 则降低，因为保持较小的车头时距是第三优先事项。尽管这些权重可针对不同驾驶模式进行配置，但为简化起见，本分析中所有车辆均使用相同的权重。MARL-MASS 策略采用学习率 $\alpha=1e^{-4}$ 进行训练，并使用五个随机种子来捕捉奖励的变化情况。在为期 20,000 个回合的训练过程中，每训练 200 个回合，就使用 20 个回合对中间策略进行评估。为训练评估中所考察的换道控制器，使用了一台共享服务器，其运行 Ubuntu 24.04，配备 AMD EPYC 9654 处理器、4×40 GB GPU（Nvidia-L40）和 1.5 TB 内存。不过，训练仅使用了约 1.5 GB GPU 内存和 3 GB RAM。

B. 安全评估

MASS在密集交通的匝道合流场景中的安全保证得到了分析；该场景极具挑战性，安全换道的机会非常少。图7绘制了训练不安全MARL、MARL-CS、MARL-HSS和MARL-MASS时观测到的最小时间车头时距。该参数的获取方式是：在训练过程中用于中间评估的各回合的每一步，从所有网联自动驾驶汽车中找出最小时间车头时距。由于不安全MARL和MARL-CS不提供安全保证，它们会做出不安全的控制决策，持续违反安全约束，从而导致碰撞。这些违规可通过HSS和MASS等安全屏障加以避免；这些安全屏障通过严格遵守0.5秒时间车头时距的安全约束来提供安全保证。

训练 MARL-HSS 时，观察到其车头间距高于 MARL-MASS，这表明道路利用率较低，且发生交通拥堵的可能性更高。为避免这一场景下出现交通拥堵，MASS 使网联自动驾驶车辆能够协同创造安全的变道机会。此外，自定义奖励函数进一步促使多智能体强化学习策略充分利用这些机会，同时保持较小的时间车头时距

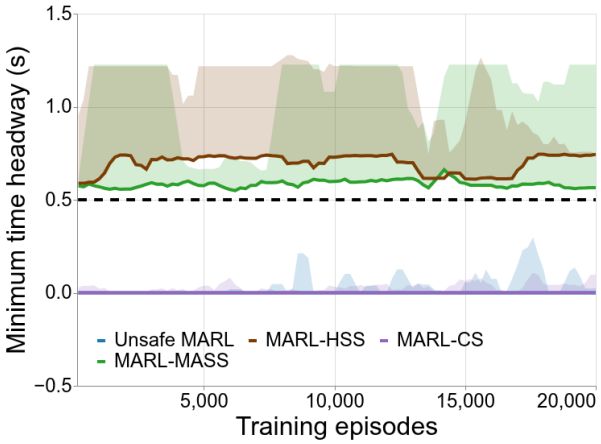


图7. 在密集交通中进行MARL训练期间观测到的最小时间车头时距。散布范围展示了在5个随机种子上观测到的最小时间车头时距的范围（最大值和最小值）。

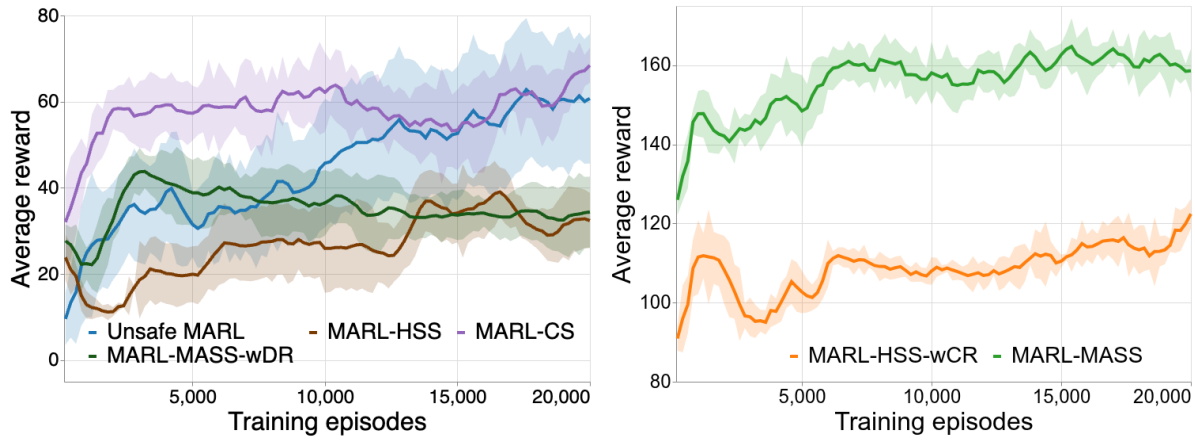
且不违反安全约束。因此，所提出的安全屏障确保严格执行基于时间车头时距定义的安全约束，即使在拥堵合流场景中遇到的复杂场景下也是如此。

*C. MARL*训练稳定性

通过比较图8中采用默认奖励和所提出的奖励函数的MARL换道控制器的奖励曲线，分析多智能体强化学习的训练稳定性。图8分为两个子图，以清晰展示自定义奖励函数和MASS的优势。图8a绘制了使用默认奖励函数（3）训练的策略的学习曲线。尽管不安全的多智能体强化学习策略在训练后期获得了与MARL-CS策略相当的奖励，但其方差更大，表明缺少集中式安全屏障时训练不稳定。正如预期，MARL-HSS策略未能收敛，因为在密集交通中很难找到安全的变道机会。采用默认奖励的MARL-MASS（记作MARL-MASS-wDR）虽然在训练初期获得了较高奖励，但最终收敛到奖励较低的策略。此外，奖励曲线的较高方差凸显了默认奖励函数无法训练出能够持续获得较高奖励的稳定MARL-MASS策略。

图 8b 展示了使用所提出的自定义奖励函数（11）训练的策略的奖励曲线，这些策略采用了集成控制屏障函数安全屏蔽器（如 HSS 或我们的 MASS）的换道控制器。在此图中，使用自定义奖励训练的 MARL-HSS策略称为 MARL-HSS-wCR。与默认奖励相比，使用所提出的自定义奖励函数的这些多智能体强化学习策略获得了更高的奖励，且方差更小。较高的奖励主要归因于速度奖励的权重 w_s 较高 r_s；而较小的方差则表明，所提出的奖励函数适用于训练配备安全屏障的换道控制器。

与 MARL-HSS 相比，所提出的控制器采用所提出的奖励函



(a) 使用默认奖励的基线控制器奖励曲线

(b) 使用定制奖励的奖励曲线

图8. 训练多智能体强化学习换道控制器的奖励曲线比较。扩散范围表示在5个随机种子上观察到的标准误差。曲线通过对5个评估轮次的窗口取平均进行平滑。

表I 使用100个拥堵汇入场景回合进行的策略评估。括号中的数字是在使用5个不同随机种子训练的策略上观察到的标准误差。

	最小 车头时距 (s)	速度（米/秒）	汇入 百分比（%）
MARL-MASS（我们的方法）	✔ **0.59(0.01)**	✔ **24.71(0.09)**	✔ **83.36(0.38)**
MARL-MASS-wDR	✔ 0.59(0.03)	↓ 23.23(0.40)	↓ 69.26(3.65)
MARL-HSS-wCR	↓ 0.64(0.02)	↓ 23.87(0.09)	↓ 69.95(1.08)
MARL-HSS	↓ 0.72(0.03)	↓ 21.95(0.31)	↓ 52.75(7.07)
MARL-CS	✖ 0.00	↑ 25.83(0.84)	↑ 89.08(2.42)
不安全MARL	✖ 0.00	↑ 26.26(1.52)	↓ 76.91(14.96)

数后获得了显著更高的奖励，如图 8b 所示。此外，MARL-MASS 策略训练期间学习曲线持续上升且方差极小，表明其收敛稳定。这些观察结果凸显了 MASS 的协作特性所带来的更多安全变道机会。

D. 多智能体强化学习策略评估

来自各换道控制器且表现最佳的策略（平均奖励最高）采用 100 个策略评估回合进行评估，这些回合不同于训练期间用于中间评估的回合。通过平均速度（m/s）、最小车头时距（s）和汇入百分比对这些策略进行比较，以分析交通效率。较高的平均速度表示较高的交通效率，而较低的最小车头时距（接近 0.5 s）则表示在遵守安全约束的同时实现了高效的道路空间利用率。汇入百分比是指在汇入道路上生成的车辆中成功汇入高速公路交通的车辆所占的百分比。由于密集交通仅提供少量安全换道机会，因此较高的汇入百分比表示更高的协同换道可能性，使更多车辆能够汇入高速公路。上述参数在策略评估回合中的平均值

价回合如表 I 所示。

在确保安全的换道控制器中，我们的换道控制器实现了大于 0.5 s 的最低时间车头时距，表明其在确保严格遵守安全约束的同时实现了高效的道路利用率。与采用默认奖励训练的 MARL-HSS 相比，使用 MASS 和自定义奖励训练 MARL（MARL-MASS）的优势，可以从中间策略评估参数的逐步改善中清晰看出；表 I 展示了采用自定义奖励的 MARL-HSS（MARL-HSS- wCR）以及采用默认奖励的 MARL-MASS（MARL- MASS-wDR）的相关结果。使用定制奖励函数训练 MARL- HSS 策略时，汇入百分比从 52.75% 小幅提升至 69.95%；而使用 MARL-MASS 时，则显著提升至 83.36%，即从 69.95% 提升。汇入百分比的增加可归因于 MASS 的协同性：它会降低自车速度，并为汇入车辆创造安全间隙。此外，平均速度提升 12.57%，进一步体现了采用 MARL-MASS 进行协同换道相较于 MARL-HSS 的优势。

相反，未提供安全保障的换道控制器（如不安全MARL和MARL-CS）被发现比我们的方法实现了略高的交通效率，平均速度也更高。尽管MARL-MASS策略的汇入百分比比不安全MARL高出 6.54%，但仍比 MARL-CS低 5.72%。然而，这些微小提升是以牺牲安全性为代价的，最小车头时距为 0.00 秒，意味着发生了碰撞。因此，所提出的换道控制器被认为能够实现与不安全基线相当的交通效率。

总之，在密集交通的匝道汇入场景中，MARL-MASS 策略能够缓解交通拥堵；与使用基线安全屏障 HSS 训练的策略相比，它们提高了交通效率。最重要的是，MARL-MASS 能够确保变道操作安全，同时交通效率可与不安全 MARL 和 MARL-CS 策略相媲美。

V. 结论

在交通拥堵的汇入路段中，网联自动驾驶车辆要在确保安全的同时高效变道，是一项艰巨的任务。本文提出了一种名为 MARL-MASS 的换道控制器，它利用多智能体强化学习提高交通效率，并通过一种新颖的协同安全屏障 MASS 提供安全保证。为促进网联自动驾驶车辆之间的协作，我们提出了一种构建交互拓扑的算法，以刻画车辆间的依赖关系。此外，对于配备安全屏障的智能体，我们定义了一种侧重于效率的自定义奖励函数。我们通过模拟密集交通下的高速公路合流场景来评估所提出的方法；在该场景中，安全汇入的机会有限。

结果表明，与基线 MARL-HSS 相比，MARL-MASS 严格遵守基于时间车头时距的安全约束，同时通过提高平均速度、增加合流比例并降低最小车头时距来提升交通效率。自定义奖励函数提高了采用安全屏障训练的多智能体强化学习策略的稳定性。尽管诸如不安全 MARL 和 MARL-CS 等基线策略实现的交通效率可与 MARL-MASS 相当，但它们无法提供安全保证。相比之下，所提出的换道控制器能够促进协作，并有效平衡相互冲突的目标，从而在拥堵合流场景中高效变道，同时确保安全。

未来的工作中，应将 MARL-MASS 扩展至其他常见交通拥堵的场景，例如双重合流和交通交叉口。此外，我们计划将 MARL-MASS 扩展至混合交通场景。

致谢

作者感谢编辑和匿名审稿人提出的宝贵意见。本出版物源自一项研究，该研究获得了 Taighde´Eireann–Research Ireland 的资助：可持续社会先进网络研究培训中心（ADVANCE CRT）的资助编号为 18/CRT/6222，CONNECT：Research Ireland 未来网络中心的资助编号为 13/RC/2077 P2。

参考文献

[1] M. Shaygan、C. Meese、W. Li、X. G. Zhao 和 M. Nejad，《利用人工智能进行交通预测：近期进展与新兴机遇综述》，《交通研究C辑：新兴技术》，2022年12月。[2] D. Schrank、L. Albert、K. Jha 和 B. Eisele，《2023年城市交通报告》，德克萨斯农工大学交通研究所，技术报告，2024年6月。[3] M. Garg 和 M. Bouroche，《网联自动驾驶车辆能否在不牺牲效率的情况下提升现实场景中的混合交通安全？》，《*IEEE*智能交通系统汇刊》，2023年6月。[4] B. Hegde 和 M. Bouroche，《网联与自动驾驶车辆中基于人工智能的变道模块设计：综述》，载于第十二届交通与运输智能体国际研讨会，维也纳，2022年。[5] S. V. Albrecht、F. Christianos 和 L. Sch¨afer，《多智能体强化学习：基础与现代方法》。麻省理工学院出版社，2024年12月。[6] C. Yu、X. Wang、X. Xu、M. Zhang、H. Ge、J. Ren、L. Sun、B. Chen 和 G. Tan，《基于动态协同图的高速公路自动驾驶分布式多智能体协同学习》，《*IEEE*智能交通系统汇刊》，2020年2月。[7] J. Dong、S. Chen、Y. Li、R. Du、A. Steinfeld 和 S. Labi，《利用深度强化学习进行空间加权信息融合：换道自动驾驶车辆战术控制与连通范围评估》，《交通研究C辑：新兴技术》，2021年7月。[8] S. Chen、J. Dong、P. Y. J. Ha、Y. Li 和 S. Labi，《用于网联自动驾驶车辆多智能体协同控制的图神经网络与强化学习》，《计算机辅助土木与基础设施工程》，2021年。[9] W. Zhou、D. Chen、J. Yan、Z. Li、H. Yin 和 W. Ge，《混合交通中网联与自动驾驶车辆协同变道的多智能体强化学习》，《自主智能系统》，2022年3月。[10] Y. Hou 和 P. Graf，《利用多智能体深度强化学习在高速公路交织区进行去中心化协同变道》，2021年10月，arXiv：2110.08124。[11] D. Chen、M. R. Hajidavalloo、Z. Li、K. Chen、Y. Wang、L. Jiang 和 Y. Wang，《混合交通中高速公路匝道汇入的深度多智能体强化学习》，《IEEE智能交通系统汇刊》，2023年。[12] B. Hegde 和 M. Bouroche，《网联与自动驾驶车辆安全变道的多智能体强化学习：综述》，《人工智能通讯》，2024年1月。[13] R. Cheng、G. Orosz、R. M. Murray 和 J. W. Burdick，《通过障碍函数实现端到端安全强化学习：面向安全关键连续控制任务》，《人工智能促进会会议论文集》，2019年7月。[14] C. Zhang、L. Dai、H. Zhang 和 Z. Wang，《控制障碍函数引导的深度强化学习：用于自动驾驶车辆匝道汇入决策》，《IEEE智能交通系统汇刊》，2025年。[15] B. Hegde 和 M. Bouroche，《利用去中心化安全屏障实现安全高效的网联自动驾驶车辆变道》，载于2025年IEEE智能车辆研讨会（IV），2025年6月。[16] D. Marinescu、J. ˇCurn、M. Bouroche 和 V. Cahill，《利用协同智能车辆实现匝道交通汇入：一种基于时隙的方法》，载于2012年第15届IEEE智能交通系统国际会议，2012年9月。[17] S. Jing、F. Hui、X. Zhao、J. Rios-Torres 和 A. J. Khattak，《用于网联自动化车辆最优汇入序列与匝道汇入控制的合作博弈方法》，《IEEE智能交通系统汇刊》，2019年11月。[18] R. Valiente、B. Toghi、R. Pedarsani 和 Y. P. Fallah，《基于强化学习的协同自动驾驶在混合自主交通中的鲁棒性与适应性》，《IEEE智能交通系统开放期刊》，第3卷，2022年。[19] J. Liang、Y. Li、G. Yin、L. Xu、Y. Lu、J. Feng、T. Shen 和 G. Cai，《基于多智能体系统的网联自动化车辆协同控制分层架构》，《IEEE车辆技术汇刊》，第72卷，2023年2月。[20] B. A. Butler、C. H. Leung 和 P. E. Par´e，《耦合多智能体系统的协作安全编队控制》，载于2024年欧洲控制会议（ECC），2024年6月。[21] X. Tan 和 D. V. Dimarogonas，《控制障碍函数在多智能体系统中的分布式实现》，《IEEE控制系统快报》，2022年。[22] B. Paden、M. ˇC´ap、S. Z. Yong、D. Yershov 和 E. Frazzoli，《城市自动驾驶车辆的运动规划与控制技术综述》，《IEEE智能车辆汇刊》，2016年3月。[23] P. Polack、F. Altch´e、B. d'Andr´ea Novel 和 A. de La Fortelle，《运动学自行车模型：一种用于为自动驾驶车辆规划可行轨迹的一致模型？》，载于2017年IEEE智能车辆研讨会（IV），2017年6月。[24] A. D. Ames、S. Coogan、M. Egerstedt、G. Notomista、K. Sreenath 和 P. Tabuada，《控制障碍函数：理论与应用》，载于2019年第18届欧洲控制会议（ECC），2019年6月。[25] S. Han、S. Zhou、J. Wang、L. Pepin、C. Ding、J. Fu 和 F. Miao，《用于网联自动驾驶车辆安全高效行为规划的多智能体强化学习方法》，《IEEE智能交通系统汇刊》，2023年。[26] E. Leurent，《用于自动驾驶决策的环境》，2018年，出版物标题：GitHub代码库。[在线]。链接：https://github.com/eleurent/highway-env