内容新鲜度保障的车联网多智能体缓存分发策略

崔亚平[1,2,3]，石宏吉[1,2,3]，吴大鹏[1,2,3]，何鹏[1,2,3]，王汝言[1,2,3]
（1.重庆邮电大学通信与信息工程学院，重庆 400065；2.先进网络与智能互联技术重庆市高校重点实验室，重庆 400065；
3.泛在感知与互联重庆市重点实验室，重庆 400065）

摘　要：车辆需要频繁动态变化内容支持车联网（IoV）时延敏感型应用，这会增加宏基站（MBS）负载，降低内容新鲜度。利用边缘缓存将最新内容提前缓存在小基站（SBS）能有效降低车辆时延和提高内容新鲜度。对影响时延和内容信息年龄（AoI）进行深入分析，提出一种内容新鲜度保障的多智能体强化学习（MARL）算法，通过优化缓存分发决策保障车辆获得高新鲜度内容。仿真结果表明，所提算法不仅收敛速度更快，而且在降低车辆时延和提升内容新鲜度方面表现出更好效果。

关键词：车联网；边缘缓存；信息年龄；多智能体强化学习
中图分类号：TN927
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025013

Multi-agent caching distribution strategy for content freshness guarantee in IoV

CUI Yaping[1,2,3], SHI Hongji[1,2,3], WU Dapeng[1,2,3], HE Peng[1,2,3], WANG Ruyan[1,2,3]
1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. Advanced Network and Intelligent Connection Technology Key Laboratory of Chongqing Education Commission of China, Chongqing 400065, China
3. Chongqing Key Laboratory of Ubiquitous Sensing and Networking, Chongqing 400065, China

Abstract: Vehicles need to dynamically changing content to support latency-sensitive applications in Internet of vehicles (IoV), thereby increasing the load on the macro base station (MBS) and reducing the freshness of content. Utilizing edge caching to cache the latest content in small base station (SBS) can effectively reduce the latency and improve the content freshness. An in-depth analysis was conducted on latency and content's age of information (AoI). A content freshness assurance multi-agent reinforcement learning (MARL) algorithm was proposed, which optimized cache distribution decisions to guarantee high freshness. Simulation results show that the proposed algorithm not only converges faster but also demonstrates better performance in reducing latency and enhancing content freshness.

Keywords: Internet of vehicles, edge caching, age of information, multi-agent reinforcement learning

---

收稿日期：2024–09–25；**修回日期：**2024–12–10
通信作者：崔亚平，cuiyp@cqupt.edu.cn
基金项目：国家自然科学基金资助项目（No.61801065, No.62271096, No.61871062, No.U20A20157, No.62061007）；重庆市教委科学技术研究基金资助项目（No.KJQN202000603, No.KJQN202300621）；重庆市自然科学基金资助项目（No.CSTB2022NSCQ-MSX0468, No.cstc2020jcyjzdxmX0024, No.cstc2021jcyjmsxmX0892, No.CSTB2023NSCQ-LZX0134）；重庆市高校创新研究群体基金资助项目（No.CXQT20017）；重邮信通青创团队支持计划基金资助项目（No.SCIE-QN-2022-04）；四川省重点研发计划基金资助项目（No.2024YFHZ0093）

Foundation Items: The National Natural Science Foundation of China (No.61801065, No.62271096, No.61871062, No.U20A20157, No.62061007), The Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJQN202000603, No.KJQN202300621), The Natural Science Foundation of Chongqing (No.CSTB2022NSCQ-MSX0468, No.cstc2020jcyjzdxmX0024, No. cstc2021jcyjmsxmX0892, No. CSTB2023NSCQ-LZX0134), The University Innovation Research Group of Chongqing (No.CxQT20017),The Youth Innovation Group Support Program of ICE Discipline of CQUPT (No.SCIE-QN-2022-04), The Key Research and Development Program of Sichuan Province (No.2024YFHZ0093)

0　引言

随着车联网（IoV, Internet of vehicles）的发展，车辆能为用户提供更广泛的高质量内容和服务，从而增强用户的驾驶体验。在车联网中，车辆可以通过车辆对一切（V2X, vehicle-to-everything）连接大量的基础设施、用户设备和智能设备，各种车辆应用快速增长，导致移动数据量急剧增加，对车联网网络负载造成极大的压力[1]。为了满足大量车辆请求不同IoV应用服务时的服务质量（QoS, quality of service），需要对车辆的计算能力、通信质量和存储资源有很高的要求；此外，在高度动态的环境中，车辆的高移动性和有限的通信距离导致连接的频繁切换，对车辆与车辆（V2V, vehicle-to-vehicle）、车辆与基础设施（V2I, vehicle-to-infrastructure）等通信的稳定连接造成影响[2-3]；最后，考虑到部分IoV应用服务内容具有时效性，但是大量的内容访问造成的网络堵塞会导致车辆具有相当大的内容交付时延，这可能会对道路行驶安全造成严重的后果。上述问题表明IoV为车辆提供应用服务时面临巨大挑战[4]。

边缘缓存作为解决上述问题的一种有效技术[5-6]，其主要思想是充分利用宏基站（MBS, macro base station）和小基站（SBS, small base station）的存储资源进行内容的缓存，在更靠近用户的网络边缘处向用户提供应用服务，这样可以减少用户和核心网之间的重复传输来缓解回程链路的压力，能有效地减少端到端的时延、提高传输效率、增强系统可靠性以及提高QoS[7]。边缘缓存主要涉及的问题是在哪里进行缓存、如何进行缓存以及缓存内容的分发问题。文献[8]提出了一种在IoV环境下基于内容请求预测的协同缓存策略，该策略将车辆请求概率较大的内容提前缓存到其他车辆或路侧单元（RSU, road side unit）中，以降低内容获取时延。首先使用聚类方法简化车辆请求与内容传输的过程，然后使用长短期记忆网络来预测来自车辆的内容请求，最后使用强化学习方法迭代更新出最优缓存决策以提高车辆请求的QoS。文献[9]设计了一个自适应的边缘缓存策略，其充分利用道路的闲置资源，构建了一个协作缓存系统，提出了一种面向IoV的社会感知的去中心化协作缓存算法，有效地减少了内容传输时延和响应时间。该算法中RSU负责全局模型的训练与更新，车辆使用本地数据向RSU提供本地更新，RSU平均所有车辆提供的更新以改进共享模型，还利用车辆用户的社交网络来获得不同地区的车辆接触率。文献[10]将内容放置和内容分发过程中的成本和时延的双目标协同优化问题建模为双时间尺度马尔可夫决策过程，并给出了不同时间尺度下的资源分配方案，有效地提高了网络效用。文献[11]在联合协作缓存和处理框架中提出了一个整数线性规划，通过确定缓存中视频流量的位置和视频请求的调度，减少了视频检索的时延成本。除此之外，车辆也可以利用自身的存储资源来缓存内容，并通过稳定可靠的V2V通信技术为车辆请求用户直接提供内容，可以进一步降低获取内容的时延[12]。文献[13]研究了IoV中内容缓存决策的优化，车辆间通过V2V链路进行协作，先是提出一种时延感知的车辆关联算法来优化车辆关联，然后根据车辆关联结果，在不同的网络场景下优化了缓存决策。

边缘缓存技术有着很多的优势，且已经在边缘缓存的部署和管理方面进行了大量的工作，但其中大多数集中在不需要频繁刷新的静态内容项上[14]。而对于随时间和环境所变化而变化的动态的缓存内容时，传统的内容放置和缓存方式可能会导致内容的时效性降低，使得边缘用户收到过时的内容，内容的价值大大降低[15]。对于车辆用户使用道路驾驶安全相关的应用服务时所需的动态缓存内容，例如实时街道地图、保障车辆安全行驶的消息等，内容的新鲜度是很重要的，过时的内容可能会导致驾驶员做出错误的行驶决策，影响驾驶安全[16]。信息年龄（AoI, age of information）被用来表征内容的新鲜度，其表示当前版本的内容自生成以来所经历的时间[17]。因此，对于动态的缓存内容应及时地将内容更新到最新的版本。然而，在频谱资源有限的情况下，缓存内容的频繁更新将导致额外的频谱资源消耗，这就导致内容分发时车辆用户的时延性能有所降低[18]。因此，当车辆用户请求进行内容服务时，不仅需要考虑如何为车辆用户的内容分发进行资源分配，以减少车辆用户时延，还需考虑基站（BS, base station）应如何对缓存内容进行更新，以保障车辆用户所获得的内容具有高的新鲜度。

对于动态变化的内容项，SBS需要及时地将内容进行更新以保证内容的新鲜度。因此，需要设计有效的缓存内容分发和资源分配的方案来优化用户

时延和内容的新鲜度。文献[19]考虑了内容推荐和AoI的时隙系统，推导出了一个整数线性规划的公式，提出了基于拉格朗日分解的高效算法来解决缓存更新的最优调度问题。当车辆用户请求信息娱乐应用服务时通过推荐系统推荐一些类似的相关内容来代替最初请求的、在该高速缓存中不存在的内容。但是对于道路安全驾驶相关的动态变化的服务内容，很难通过寻找类似的相关内容来代替。文献[20]考虑了在车辆网络中为了保证车辆接收的内容的新鲜度，提出了一种缓存辅助的时延更新和分发方案，首先推导了车辆接收的内容的AoI和服务时延的闭环公式，然后对内容更新、分发和无线电资源分配进行联合优化，以满足不同应用的多样化服务时延和AoI要求。进一步地，文献[21]考虑在信息为中心的车辆网络（ICVN, information-centric vehicular network）中为车辆用户提供动态的与驾驶相关的上下文信息，提出了2种不同的方案来进行路边单元的缓存更新和内容分发，并分别对AoI和服务时延进行了权衡分析。文献[22]研究了在BS存储容量有限的情况下对内容缓存的最优调度，通过最小化与内容下载、内容更新和AoI成本相关的惩罚函数来减轻回程链路的负载。文献[23]考虑到边缘节点处理容量的有限性，提出了一种联合考虑信息新鲜度与调度实时性的调度方法。该方法首先利用队列的系统时间和信息年龄分别刻画任务在计算之前的时延和计算之后的信息新鲜度，同时给每个卸载任务设定合理的截止期限，来保证任务进入计算过程之前的有效性。文献[24]提出了一种联合优化的卸载与调度策略。该策略通过选择在传感节点或Sink处理数据，最小化AoI与能耗的加权和。

以上研究提供了针对不同应用场景和网络条件的内容缓存更新和分发优化方案，通过多种方法和技术有效地提高了内容的新鲜度和用户服务质量。然而，这些方法在面对复杂的动态环境时往往缺乏灵活性，特别是在不完全信息的条件下，难以做出及时有效的决策。强化学习（RL, reinforcement learning）是解决缓存内容的分发和资源分配的一种有效方案，其通过与动态变化的环境进行交互，能够在不完全信息下快速决策[25]。已经有很多研究基于RL的方法设计出了缓存内容分发策略的方案来满足延时敏感型应用的服务时延和AoI要求。文献[26]关注需要及时更新以确保其相关性的动态内容的缓存，利用强制分解技术和深度强化学习，提出了一个基于约束马尔可夫决策过程的用户请求队列感知的缓存内容更新调度算法，以最小化分发给用户的动态内容的平均AoI。文献[27]制定了一个动态的状态更新优化问题，同时考虑了用户的AoI和能源消耗，提出了一种无模型的强化学习算法，以最小化长期累积成本的期望。文献[28]将边缘节点当作智能体，把缓存更新问题建模为一个协作的多智能体马尔可夫决策过程，设计了一种离散的多智能体强化学习（MARL, multi-agent reinforcement learning）算法，每个边缘节点只基于本地观测做出决策，以最小化长期平均加权成本。

上述的MARL算法中智能体只是基于本地的观测进行决策，没有考虑相邻智能体之间的相互影响，在真实环境的系统中的各个SBS之间是相互影响相互制约的，并不是独立存在的。因此，本文提出一种内容新鲜度保障的MARL算法对车辆用户进行内容分发决策和资源分配决策。在MARL算法中加入注意力机制，将内容分发决策和资源分配决策问题转化为部分可观测马尔可夫决策过程（POMDP, partially observed Markov decision process），并充分利用智能体之间的通信交换信息，使得每个智能体在做决策时评估相邻智能体对本身的影响大小，从而优化决策，降低车辆用户所获得内容的AoI和用户时延，提高网络的效用，更好地适应环境的变化。

本文的主要贡献如下。

1) 为了支持驾驶安全相关的应用服务，考虑对动态的内容进行边缘缓存。首先，在内容更新时，采用新鲜度感知的高速缓存刷新方案；在内容分发时，考虑3种不同的内容分发方式。然后，将用户请求服务建模为智能决策优化问题，以保障车辆用户获得高新鲜度和低时延的内容。

2) 提出了内容新鲜度保障的MARL算法进行内容的分发策略和资源分配策略。首先将建立的优化问题转化为POMDP，把每个SBS当作智能体；然后智能体通过注意力机制来评估相邻智能体的影响，进而优化缓存分发决策，更好地适应环境的变化。

3) 通过执行内容新鲜度保障的MARL算法，SBS智能体实时做出内容分发策略和资源分配策

略。与基准算法相比，证明了所提算法的有效性，其具有更快的收敛性能，并且在用户时延和保障内容新鲜度方面也有更好的性能。

1　系统模型

本文考虑由一个MBS覆盖多个SBS以及多个车辆组成的异构车辆网络，如图1所示。SBS一方面从MBS处获取内容项进行缓存，另一方面将缓存的内容项向提出服务请求的车辆分发，所有车辆用户的请求率λ服从泊松分布。

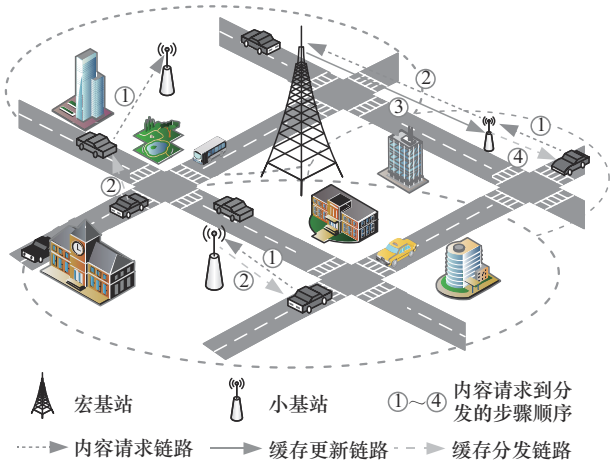


图1　系统模型

1.1　网络模型

考虑整个场景下的总带宽为B，假设SBS从MBS获取内容所使用的带宽为$B_R=\theta B$，SBS进行内容分发所使用的带宽为$B_D=(1-\theta)B$，带宽比$\theta\in(0,1)$。MBS的覆盖范围内有M个SBS，表示为$\mathcal{M}=\{1,\cdots,m,\cdots,M\}$，第$m$个SBS范围内有$N$个车辆用户，表示为$\mathcal{N}_m=\{1,2,\cdots,i,\cdots,N\},m\in\mathcal{M}$。在MBS的服务范围内，所有的车辆用户被允许使用V2I通信接入MBS覆盖下的SBS来获取缓存内容，c_{v}、c_{SBS}和c_{MBS}分别表示车辆、SBS与MBS各自的处理数据能力，即每秒的CPU运行周期数。另外网联车辆也被允许使用V2V通信来从其他网联车辆处获取缓存内容[29]。

1) V2I通信

在车辆网络中，将时间用时隙$\mathcal{T}=\{0,1,\cdots,t,\cdots,T\}$来表示，每个时隙的间隔时间可能是几毫秒。根据香农公式[30]，车辆i与SBS_m通过连接V2I链路进行通信时的传输速率为

$$r_{i,m}^{t}=B_{m,i}\mathrm{lb}\left(1+\frac{P_mG_{i,m}^{t}}{\sigma^2}\right),m\in\mathcal{M},i\in\mathcal{N}_m \tag{1}$$

其中，P_m为SBS_m的发送功率，$B_{m,i}\in B_D$为SBS_m为车辆i所分配的通信带宽，σ^2为加性白高斯噪声的功率谱密度，$G_{i,m}^{t}=d_{m,i}^{-\alpha}$为车辆$i$与$\mathrm{SBS}_m$之间的信道增益，其随车辆$i$与SBS之间的距离$d_{m,i}$变化而变化，$\alpha$为路径损耗常数。

2) V2V通信

为了最大化频谱利用率，车辆i与车辆j之间的V2V通信链路使用的频谱资源$B_{i,j}\in B_m$，为复用SBS_m内的一个子带频谱，车辆i与车辆j之间的传输速率为

$$r_{j,i}^{t}=B_{i,j}\mathrm{lb}\left(1+\frac{P_jG_{i,j}^{t}}{I_{\mathrm{Intf}}+\sigma^2}\right),i,j\in\mathcal{N}_m \tag{2}$$

其中，I_{Intf}是干扰功率，P_j为车辆j的固定发送功率，$G_{i,j}^{t}$是车辆i与车辆j之间V2V链路的信道增益。

1.2　内容分发模型

为了简化，将MBS看作一个提供动态信息的内容服务器。当车辆用户发送内容请求时，为减轻MBS通信负载，MBS可以通过无线传输的方式将内容项提前缓存于SBS中，由SBS将缓存内容分发给车辆用户。

在SBS处采用新鲜度感知的高速缓存刷新方案[31]。具体地，为内容项设置AoI刷新窗口阈值以保证内容的新鲜度。SBS通过单信道以先进先出（FIFO, first-in-first-out）的方式为车辆用户提供服务。当车辆用户提出对内容项的请求时，如果信道未被占用且AoI低于刷新窗口阈值，则立即向车辆用户服务，否则在队列中等待。SBS在向车辆用户分发内容前始终监测内容项的新鲜度，如果内容项的AoI大于刷新窗口阈值，SBS则向MBS获取内容项的最新版本，刷新缓存，然后将其分发给车辆用户。本文假设整个系统用于内容更新和内容分发的上下行频谱资源是固定的。频繁的更新缓存内容会使得平均AoI变小，但也会导致用户时延的增加，而过低频率更新缓存内容会使得系统平均AoI增大，用户时延减小，因此刷新窗口阈值的设置存在对系统AoI和用户时延之间存在一定的权衡。SBS处缓存内容的AoI及刷新窗口阈值W如图2所示。

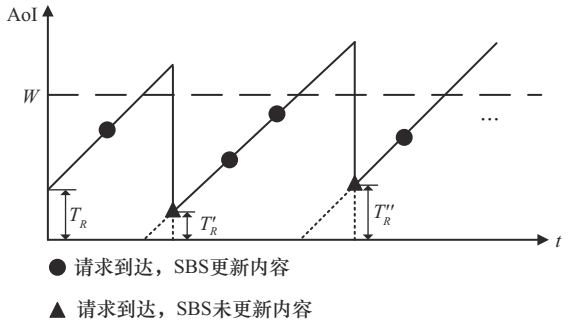


图2　SBS内容更新的AoI

如图2所示，SBS处的AoI刷新窗口阈值为W；T_R表示SBS与MBS通信更新内容所需的时间，T_R的大小与分配的带宽θB呈负相关。另外，忽略了MBS获取道路设备传感器数据的传输时间，假设内容数据在MBS处产生[21]。因此，用A_t表示在时间t时SBS处内容的AoI，其计算方法如式(3)所示。

$$A_t=\begin{cases}A_{t-1}+1, & \text{SBS处内容未更新}\\ T_R, & \text{SBS处内容更新}\end{cases} \tag{3}$$

当车辆i内容请求发出后，MBS可以通过3种内容分发方式进行决策：第一种是车辆用户与邻近车辆通过V2V链路进行内容的交付；第二种是车辆与SBS通过V2I链路直接进行内容的交付，不需要进行内容更新；第三种是SBS先与MBS通信，从MBS处获取最新的内容，然后再与车辆通信交付内容。

1) 请求车辆i从邻近的目标车辆j直接进行内容交付。只有当请求车辆与邻近车辆的距离$d_{i,j}$小于允许V2V通信的最大距离$d_{\max}$，以及邻近的目标车辆处内容的AoI小于刷新窗口阈值时，请求车辆才能通过V2V链路通信从邻近的目标车辆j获得内容，用β_i来表示车辆用户i是否满足V2V内容分发条件。$\beta_i=1$即表示满足V2V内容分发条件；反之，$\beta_i=0$。当$\beta_i=1$时，车辆用户从发起服务请求到服务完成的用户时延以及AoI可以表示为

$$D^t_{j,i}=\frac{S}{r_{j,i}}+\frac{C}{c_{\mathrm{v}}} \tag{4}$$

$$A^t_{j,i}=A_j+D^t_{j,i} \tag{5}$$

其中，S表示服务内容的数据包大小，C表示缓存内容的分发任务所需的CPU执行周期数，A_j表示通过V2V链路进行内容交付时，邻近车辆处内容的AoI，$A_j<W$。

2) 车辆i从SBS_m直接进行内容交付。当车辆请求到达时，SBS处的AoI小于刷新窗口阈值W，则SBS直接向车辆分发内容。φ_i表示SBS处内容AoI是否满足最大刷新窗口阈值限定。$\varphi_i=1$表示SBS处内容AoI小于最大刷新窗口阈值，可直接向车辆用户交付内容；反之，$\varphi_i=0$表示SBS先更新内容，然后向车辆用户交付内容。当$\varphi_i=1$时，车辆用户从发起服务请求到服务完成的用户时延以及AoI可以表示为

$$D^t_{m,i}=\frac{S}{r_{m,i}}+\frac{C}{c_{\mathrm{SBS}}} \tag{6}$$

$$A^t_{m,i}=A_{\mathrm{RSU}}+D^t_{m,i} \tag{7}$$

其中，A_{RSU}表示车辆i从SBS_m直接进行内容交付时SBS_m处内容的AoI，可由式(3)得到。

3)SBS_m与MBS通信，从MBS处获取最新的内容，然后再与车辆i通信进行内容分发，SBS_m与MBS通信所需的频谱带宽为B_R。当$\varphi_i=0$时，车辆用户从发起服务请求到服务完成的用户时延以及AoI可以表示为

$$D^t_{m,i}=T_R+\frac{S}{r_{m,i}}+\frac{C}{c_{\mathrm{MBS}}}+\frac{C}{c_{\mathrm{SBS}}} \tag{8}$$

$$A^t_{m,i}=T_R+\frac{S}{r_{m,i}}+\frac{C}{c_{\mathrm{MBS}}}+\frac{C}{c_{\mathrm{SBS}}} \tag{9}$$

由于SBS先更新内容再进行分发，且假设的MBS处内容项的AoI为0，因此此时车辆用户获得内容的AoI和用户时延是相等的。

1.3　模型的适用范围和限制

本文考虑的网络模型和内容分发模型适用于城市等其他交通密集区域，在这些区域内，SBS可以充分覆盖并为大量车辆用户提供服务。以及基础设施较为固定且资源配置明确的环境，如城市道路或特定的高速公路路段，尤其适合需要频繁更新且内容新鲜度要求高的场景，如实时交通信息或应急响应。然而，在车辆移动速度快、基站覆盖不足的环境下，网络模型和内容分发模型的适用性可能受到限制。此外，内容分发模型假设频谱资源固定且SBS缓存容量充足，这简化了研究模型复杂性，使其能够专注于内容分发策略和新鲜度管理。

考虑到SBS缓存容量在实际部署中可能存在限制，本文针对缓存容量不足的场景，引入了内容优先级排序和缓存替换策略，以提升缓存资源的利用效率并保障内容新鲜度。为了在有限缓存空间内优先存储对用户更有价值的内容，本文基于内容请求频率和AoI对内容进行优先级排序。请求频率越高

的内容被认为具有更大的访问需求，而信息年龄越低的内容则能更好地满足车辆用户对新鲜信息的需求。在缓存容量受限且需要腾出空间时，为保障内容新鲜度和服务质量，本文提出了基于低优先级的替换（LPR, lowest priority replacement）策略。该策略优先替换优先级最低的内容，确保缓存中的内容始终对用户需求具有较高价值。与传统的最少最近使用（LRU, least recently used）策略相比，LPR策略在动态环境中对内容价值的评估更加精准，能够更好地平衡内容新鲜度与请求频率。

2　问题描述

通过对3种内容分发策略下的AoI和服务时延公式分析，可以得出3种内容分发策略下用户的服务时延以及AoI都不相同，同时3种分发策略所消耗的资源也不相同。

本文的目标是在有限的频谱资源的情况下，在一定时间 T 内最小化整个MBS覆盖范围下所有车辆用户请求的服务时延和AoI，即找到最优的内容分发策略使整个宏基站内所有车辆用户请求的服务时延和AoI最小，问题公式建模为

$$\begin{aligned}&\min \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{i=1}^{q_m^t}a_i^m\left(A_{m,i}^t+D_{m,i}^t\right)\\&\text{s.t. C1:}a_i^m\in\{0,1,2\},\forall i\in\mathcal{N}_m,\forall m\in\mathcal{M}\\&\quad\text{C2:}\sum_{m=1}^{M}\sum_{i=1}^{q_m^t}B_{m,i}^t\leqslant B_D,\forall i\in\mathcal{N}_m,\forall m\in\mathcal{M}\\&\quad\text{C3:}r_i^t\geqslant r_{\min},\forall i\in\mathcal{N}_m,\forall t\in\mathcal{T}\end{aligned}\tag{10}$$

其中，q_m^t 表示在 t 时刻 SBS_m 范围内车辆用户请求的数量，$\boldsymbol{D}^t=(a_{m,i}^t{:}i\in\mathcal{N}_m,m\in\mathcal{M})$ 为内容分发策略矩阵，$\boldsymbol{B}^t=(B_{m,i}^t{:}i\in\mathcal{N}_m,m\in\mathcal{M})$ 为频谱资源分配矩阵。限制条件C1表示SBS向车辆用户分发内容的3种分发决策；限制条件C2表示为所有的车辆分配的频谱资源应小于用于内容分发的总频谱带宽 B_D；限制条件C3表示所有建立的通信链路应大于允许通信的最小速率。

为了找到一个最优的内容分发策略使整个宏基站内所有车辆用户请求的服务时延和AoI最小，需要尝试不同的分发策略，并对每个策略进行评估。这涉及对不同策略的选择和优化问题的搜索，这个过程可能需要时间去尝试大量的可能性，因此这是一个NP问题。传统解决NP问题的方法（迭代算法和动态规划）可能会面临指数级的时间复杂性，导致计算变得非常耗时和低效。然而，内容新鲜度保障的MARL算法通过将注意力机制加入MARL框架中，与环境进行有效的交互具有解决这个问题的潜力。

3　多智能体协作边缘缓存分发

为了解决上述问题，本节将通过内容新鲜度保障的MARL算法做出内容分发策略及资源分配策略。具体地，SBS作为智能体通过车辆用户服务请求、SBS处内容AoI大小情况做出内容分发的决策，并为内容分发的链路进行频谱资源的分配，在有限资源成本上以最小化车辆用户服务时延与AoI。

3.1　多智能体强化学习

在多智能体强化学习框架中，每个智能体都能独立地进行决策，根据环境的反馈和奖励来学习，通过不断迭代优化网络参数来改进其策略。多智能体强化学习的目标是找到一种最优解的策略，使得每个智能体的行为最大程度地促进整体系统的性能和效益[32]。

为了解决内容分发策略及资源分配问题，将问题转化为POMDP，具体地，将每个SBS当作一个智能体，其整个多智能体强化学习框架下的状态、动作及奖励的定义如下。

1) 全局状态：整体系统的状态为 M 个SBS的自我状态感知及其可以观测到的车辆环境，因此，在 T 个时隙内的状态空间可以表示为

$$\mathcal{S}(t)=\left\{s_1(t),\cdots,s_m(t),\cdots,s_M(t)\right\}\tag{11}$$

其中，$s_m(t)$ 表示 SBS_m 在 t 时隙下的状态，可以表示为

$$s_m(t)=\left\{A_m^t,B_m^t,V_m(t)\right\}\tag{12}$$

其中，A_m^t 和 B_m^t 分别表示 SBS_m 在 t 时隙下内容项的AoI大小和可用频谱资源大小，$V_m(t)=\left\{\left\{x_i,y_i,\beta_i\right\}_{i\in q_m^t}\right\}$ 表示 SBS_m 在 t 时隙下观测到车辆环境信息，$\left\{x_i,y_i,\beta_i\right\}$ 分别表示请求车辆 i 的横纵坐标及是否满足V2V通信链路的条件。

2) 观测：因为考虑的是PODMP，每个智能体并不能观测到整个环境的状态，只能获取自己覆盖范围的观测，每个智能体的观测可以表示为

$$o_m(t)=s_m(t)\tag{13}$$

3) 动作：每个SBS的actor网络根据策略 π_{η_m} 从动作空间中选择动作，即 $\pi_{\eta_m}\left(a_m|o_m\right)$，动作空间可以表示为

$$\mathcal{A}(t)=\left\{a_1(t),\cdots,a_m(t),\cdots,a_M(t)\right\} \tag{14}$$

其中，$a_m(t)=\left\{\{a_{m,i},B_{m,i}\}_{i\in q_m^t}\right\}$表示$\mathrm{SBS}_m$在$t$时隙下的内容分发方式及资源分配。

4) 奖励：奖励能反映出在一定状态下动作的好坏，这里将目标函数中的AoI和服务时延作为奖励，因为多智能体强化学习是最大化奖励值，所以在t时隙下的奖励可以表示为

$$r_t\left[\mathcal{S}(t),\mathcal{A}(t)\right]=-\sum_{m=1}^{M}\sum_{i=1}^{q_m^t}\left(A_{m,i}^t+D_{m,i}^t\right) \tag{15}$$

SBS智能体的目标就是迭代优化学习一个能最大化累计奖励的策略函数π_η，通过策略函数π_η获得累计奖励的期望表示。

$$\mathcal{R}=E_{\mathcal{A}\pi^*}\left[\sum_{t=0}^{\infty}\gamma^t r_t\left(\mathcal{S}(t),\mathcal{A}(t)\right)\right] \tag{16}$$

其中，$\gamma\in[0,1]$是折扣因子，它决定了该策略在多大程度上倾向于短期回报而不是长期收益。

3.2　内容新鲜度保障的MARL

在多智能体强化学习框架下，每个SBS智能体都有自己的演员（actor）网络和评论家（critic）网络，当智能体SBS_m actor网络和critic网络进行网络更新时，所收集的经验数据（例如，状态、动作和奖励）都是智能体SBS_m本地的数据，并未考虑其相邻的其他SBS的状态、动作或奖励对SBS_m决策的影响。另外，SBS_m相邻的其他SBS对其决策影响的重要性也不一样。

内容新鲜度保障的MARL算法通过在MARL算法框架中引入注意力机制，使得每个SBS智能体不仅可以利用相邻SBS的策略信息进行学习，以帮助建立其自己的缓存分发策略；还能在相邻SBS中动态性地选择交互的目标，调整不同目标信息的重要性和关注度，以期望找到对自己决策有益的信息进行学习，从而减少冗余信息的传递，使SBS智能体能够更好更快地学习到有效的策略[33]，保障车辆用户获得高新鲜度的内容。总体来说，内容新鲜度保障的多智能体强化学习通过引入注意力机制和灵活的信息交流方式，提供了更精确、灵活和高效的决策和学习能力。这使得智能体能够更好地适应复杂的任务和环境，并提高系统的性能和学习效率。

内容新鲜度保障的MARL框架如图3所示，其主要思想是每一个SBS智能体通过选择性地关注其他智能体的信息来学习自己的critic网络[34]。具体的，智能体SBS_m向其他智能体查询关于它们的观测和动作的信息，并将该信息合并到其自身的动作价值函数$Q_m(\mathcal{S},\mathcal{A})$。

SBS_m的观测值o_m和a_m可以被多层感知器（MLP, multi-layer perceptron）嵌入函数g_m嵌入并转换为嵌入向量$\boldsymbol{e}_m$，即$\boldsymbol{e}_m=g_m(o_m,a_m)$，那么$\boldsymbol{e}=\left\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n,\cdots,\boldsymbol{e}_M\right\}$则定义为智能体$\mathrm{SBS}_m$相邻的SBS的嵌入向量。智能体$\mathrm{SBS}_m$的critic网络具有注意力机制的动作价值函数可以表示为

$$Q_{\zeta_m}(\mathcal{S},\mathcal{A})=f_m(\boldsymbol{e}_m,\boldsymbol{x}_m) \tag{17}$$

其中，f_m是两层MLP函数，ζ_m是智能体SBS_m的目标

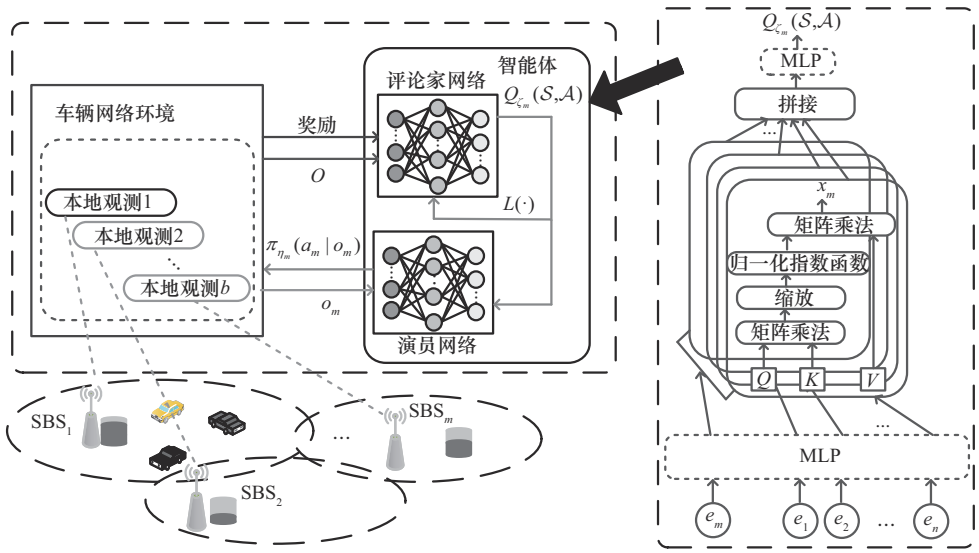


图3　内容新鲜度保障的MARL框架

critic网络的参数，$\boldsymbol{x}_m$是其他智能体对智能体SBS_m的影响，是每个智能体价值的加权和，可以表示为

$$x_m = \sum_{n \neq m} \boldsymbol{\alpha}_n \boldsymbol{v}_n = \sum_{n \neq m} \boldsymbol{\alpha}_n h(\boldsymbol{V}\boldsymbol{e}_n) \tag{18}$$

其中，$\boldsymbol{v}_n$是通过一系列变换得到的，先用除智能体m以外其他智能体的所有动作和状态用嵌入函数编码，然后通过共享矩阵$\boldsymbol{V}$线性变换，最后通过$h(\cdot)$非线性转换函数变换得到；注意力权重$\boldsymbol{\alpha}_n$是通过双线性映射比较$\boldsymbol{e}_m$和$\boldsymbol{e}_n$之间的相似性，然后通过softmax层对结果值进行归一化生成的，可以表示为

$$\boldsymbol{\alpha}_n = \frac{\exp\left(\boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{q}_m^{\mathrm{T}} \boldsymbol{k}_n \boldsymbol{e}_n\right)}{\sum\limits_{k \neq m} \exp\left(\boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{q}_m^{\mathrm{T}} \boldsymbol{k}_k \boldsymbol{e}_k\right)} \tag{19}$$

其中，$\boldsymbol{q}_m$、$\boldsymbol{k}_n$、$\boldsymbol{k}_k$是通过嵌入向量$\boldsymbol{e}_m$通过线性变换得到的。每个智能体SBS的critic网络的查询是独立的，所以这并不是一个集中式的critic网络。

在注意力机制下的智能体SBS_m的关注型critic网络可以通过计算优势函数来获取当前动作相对于其他动作的优势，并将其应用于自身的策略梯度更新中，优势函数可以表示为

$$\mathrm{Att}_m^{\mathcal{S}}(\mathcal{S},\mathcal{A}) = Q_{\zeta_m}(\mathcal{S},\mathcal{A}) - \sum_{a'_m \in \mathcal{A}} \pi_{\eta_m}\left(a'_m | o_m\right) Q_{\zeta_m}\left(\mathcal{S},\left(a'_m, a_{\setminus i}\right)\right) \tag{20}$$

其中，$a_{\setminus i}$表示除第i个智能体以外的所有智能体的动作。为了计算智能体SBS_m的梯度估计，从所有智能体的当前策略中对所有的动作空间$\mathcal{A}$抽样K个样本，这样做能避免从经验回放池里进行抽样的，导致过度泛化，使得智能体无法根据当前策略进行协调。内容新鲜度保障的多智能体强化学习算法的策略梯度更新公式可以表示为

$$\nabla_{\theta_m} J\left(\pi_{\theta_m}\right) \approx \frac{1}{K} \sum_{i=1}^{K} \left[\nabla_\theta \ln\left(\pi_{\eta_m}\left(a_m | o_m\right)\right) \mathrm{Att}_m\left(\mathcal{S},\mathcal{A},\zeta_m\right)\right] \tag{21}$$

由于各智能体SBS的参数是共享的，它们各自的策略梯度更新都由同一个损失函数，所有的critic网络一起更新以最小化联合损失函数，损失函数可以表示为

$$L\left(\zeta_m\right) = \frac{1}{K} \sum_{i=1}^{K} \left\{\gamma\left[Q_{\bar{\zeta}_m}(\mathcal{S}',\mathcal{A}') - \rho \ln\left(\pi_{\bar{\theta}_m}\left(a'_m | o'_m\right)\right)\right] + r_{\mathcal{S},\mathcal{A}} - Q_{\zeta_m}(\mathcal{S},\mathcal{A})\right\}^2 \tag{22}$$

其中，$\gamma \in [0,1]$是折扣因子，决定了策略对即时奖励与长期收益的偏好程度；$\bar{\zeta}_m$和$\bar{\theta}_m$分别为critic网络和Target策略的网络参数；$r_{\mathcal{S},\mathcal{A}}$表示即时奖励，它是智能体在状态$\mathcal{S}$下选择动作$\mathcal{A}$后得到的即时回报；$\rho$是确定熵最大化和奖励之间的平衡的温度参数[31]。内容新鲜度保障的MARL算法如算法1所示。

算法1　内容新鲜度保障的MARL算法

输入　M个智能体的actor网络参数、critic网络参数及其M个并行环境

输出　内容分发策略，频谱资源分配策略

1) for $t = 1,\cdots,T$ do
2)　每个智能体获取当前观测状态o_m^t
3)　actor网络根据策略π_{θ_m}从动作空间中选择动作$a_m^t \sim \pi_{\theta_m}(\cdot | o_m^t)$
4)　每个智能体执行动作a_m^t，与环境交互得到新的观测状态o_m^{t+1}，获得奖励r_t
5)　每个智能体将当前动作a_m^t、观测状态o_m^{t+1}与邻近智能体进行信息共享
6)　$T_{\mathrm{update}} = T_{\mathrm{update}} + 1$
7)　if $T_{\mathrm{update}} \geqslant$更新的最小批次 then
8)　 for 智能体m do
9)　　将观测值o_m和a_m转换为嵌入向量$\boldsymbol{e}_m$
10)　　计算其他智能体的对智能体SBS_m的影响x_m
11)　　根据式(17)计算$Q^{\zeta_m}(\mathcal{S},\mathcal{A})$
12)　　根据target策略计算$a'_m = \pi_{\bar{\theta}_m}(\cdot | o'_m)$
13)　　根据式(17)计算$Q^{\bar{\zeta}_m}(\mathcal{S}',\mathcal{A}')$
14)　　计算当前动作相对于其他动作的优势$\mathrm{Att}_m^{\mathcal{S}}(\mathcal{S},\mathcal{A})$
15)　　对所有的动作空间$\mathcal{A}$抽样K个样本
16)　　根据式(21)更新策略梯度$\nabla_{\theta_m} J\left(\pi_{\theta_m}\right)$
17)　　根据式(22)更新critic网络损失函数$\nabla L\left(\zeta_m\right)$
18)　　更新网络参数：$\bar{\zeta} = \tau\bar{\zeta} + (1-\tau)\zeta$，$\bar{\theta} = \tau\bar{\theta} + (1-\tau)\theta$
19)　 end for
20)　$T_{\mathrm{update}} = 0$
21)　end if
22) end for

算法1中的内容新鲜度保障的MARL算法是智能体之间通过协同做出内容缓存分发策略及资源分配策略，以此来达到车辆用户获得高新鲜度内容的目的。首先，算法输入多个智能体的actor网络参数和critic网络参数，并在并行环境中初始化智能体的网络参数及车辆环境。随后，智能体在每个时间步中获取当前的观测状态，actor网络根据当前策略从动作空间中选择最优动作，执行该动作后与环境交互，获取新的观测状态和奖励。接着，智能体将当前的动作和观测状态与邻近智能体进行信息共享，以增强决策的协同性，减少冗余信息的传递。在执行了一定数量后，算法进入网络更新阶段。智能体通过计算嵌入向量$Q_{\zeta_m}(\mathcal{S},\mathcal{A})$处理观测值和动作，并使用注意力机制评估其他智能体对自身决策的影响$\boldsymbol{x}_m$。通过计算优势函数$\mathrm{Att}_m^S(\mathcal{S},\mathcal{A})$来获取当前动作相对于其他动作的优势。随后，从所有智能体的当前策略中对所有的动作空间$\mathcal{A}$抽样K个样本，避免模型的过度泛化。智能体m计算critic网络的损失函数、通过策略梯度公式调整actor网络参数优化整体策略。当算法达到设定的最大时间步数或系统收敛时，训练终止，并输出最优的内容分发策略和频谱资源分配策略。此外，算法还根据车辆用户请求率的变化动态调整AoI刷新窗口阈值和带宽分配比，以适应不同的网络环境，优化系统性能。在整个车辆环境中，当车辆用户请求率λ比较大时，为了保障车辆用户接收到所能容忍的最低新鲜度要求的内容，AoI的刷新窗口阈值W与带宽分配比θ应为最低新鲜度服务要求的固定值，此时用户时延和系统的平均AoI主要通过内容分发策略以及资源分配策略来最小化AoI和服务时延；当车辆用户请求率λ比较小时，还可以通过先减小刷新窗口阈值W和增加带宽占比值θ来降低AoI，把更多的资源用来内容的更新，然后再进行内容分发策略以及资源分配策略来最小化AoI和服务时延。

关键参数的设置对算法性能有显著影响。探索率控制智能体在探索和利用之间的平衡，初期较高的探索率有助于智能体充分了解环境，避免早期收敛到局部最优解。折扣因子用于平衡短期和长期奖励，较高的折扣因子鼓励智能体关注长期收益，使其在复杂任务中能够更好地规划行动策略。软更新速率控制目标网络的参数更新，使得更新过程更平滑，有助于稳定训练。actor网络和critic网络的学习率设置影响训练的速度和稳定性，较低的学习率可以避免模型训练中的振荡，使模型更稳定地收敛到最优解。通过这些关键参数的合理设置，算法可以在探索与利用、短期与长期收益以及训练稳定性之间找到最佳平衡，从而实现最优性能。

4 仿真分析

4.1 结果分析

在仿真中，设置BS的带宽为10 MHz，MBS的覆盖范围为400 m，并在此范围内部署2个SBS。仿真环境的设置参考了3GPP[35]的规范，设置需要缓存的数据大小S为[5,10] MB。具体的仿真参数设置如表1所示[36]。所有仿真均在Anaconda 3和PyCharm 3.6环境下进行，算法在AMD i7-6800H CPU和NVIDIA RTX 3060 GPU上运行。

表1 仿真参数设置

参数	值
探索率	0.6
折扣因子	0.75
软更新速率	0.01
Actor网络的学习率	0.000 1
Critic网络的学习率	0.001
MBS数量	1
SBS数量	4
系统总带宽/MHz	10
内容数据大小/MB	[5, 10]
MBS发送功率/dBm	35
SBS发送功率/dBm	33
车辆发送功率/dBm	30
噪声功率/dBm	−114

通过设置仿真初始值，构建一个符合实验需求的仿真环境。通过运行仿真程序，生成用于评估的仿真数据集。基于这些数据，应用内容新鲜度保障的MARL算法进行内容分发和资源分配的优化。测试指标包括系统奖励、平均Q值、AoI、用户时延以及算法收敛性。这种方法确保了实验的可控性和一致性，为准确评估内容新鲜度保障的MARL

算法的有效性提供了坚实的基础。

在仿真中，除了所提算法外，对比算法包括多智能体深度确定性策略梯度（MADDPG, multi-agent deep deterministic policy gradient）算法[37]、双时延深度确定性策略梯度（TD3, twin delayed deep deterministic policy gradient）算法[38]和随机算法。MADDPG 算法能够有效应对多智能体环境中的连续动作空间问题，适合于本文中的 SBS 协作场景。TD3 算法通过引入 2 个独立的价值网络来减少 Q 值估计的偏差，从而提高算法的稳定性，在许多连续控制任务中表现出色，作为对比算法可以进一步验证 MARL 算法在复杂环境中的表现和优势。随机算法被选作对比基准，主要是为了提供一个没有学习和优化能力的基础参考。这种算法不依赖环境状态做出决策，而是随机选择动作，能够展示在没有任何智能决策的情况下系统的性能表现。通过将随机算法与强化学习算法进行对比，可以更直观地评估这些智能算法在优化内容分发和资源分配方面的优势，从而验证所提算法的有效性。通过这些对比，可以全面评估所提算法在优化内容分发和资源分配方面的潜力。

首先验证了算法的总体收敛性。图 4 为所提算法、MADDPG 算法和 TD3 算法经过多次迭代后达到收敛的系统奖励。

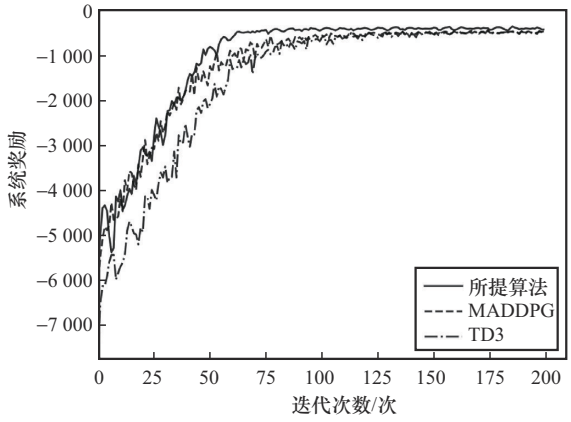
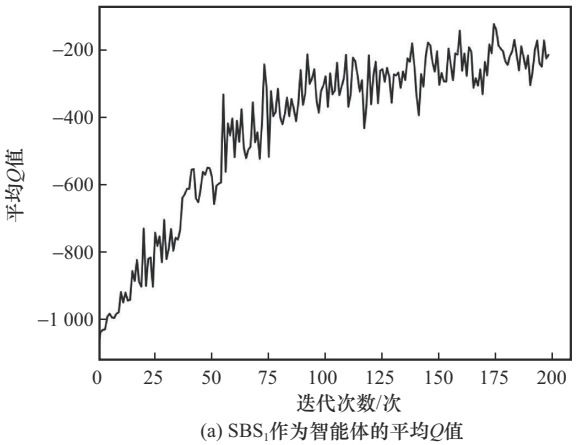
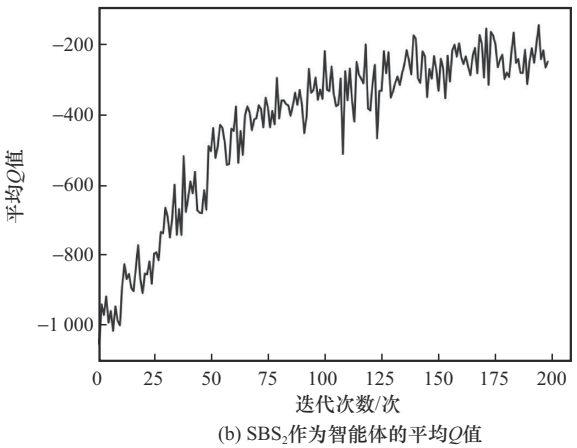


图4　3 种算法经过多次迭代后达到收敛的系统奖励

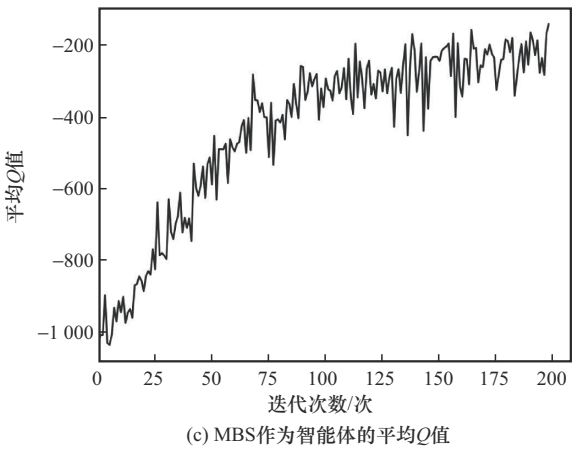
从图 4 可以看到，随着迭代次数的增加，3 种算法系统奖励呈逐渐增大的趋势，最后达到收敛。这说明 3 种算法都能通过学习经验获得知识，从而选择最优的策略动作。另外，从图 4 还可以看到，所提算法比另外 2 种基准算法有更快的收敛速度，提前了 25 次迭代达到收敛；并且所提算法与基准算法相比，系统奖励值更大，性能更好。图 5 为不同智能体收敛的平均 Q 值。所提算法设置同样的全局奖励作为每个智能体的奖励，因此智能体收敛的趋势及速度相差不大，收敛的曲线十分相似。与图 4 相类似，所有智能体在迭代 75 次之前奖励的趋势变化比较明显，之后逐渐达到收敛。



(a) SBS_1作为智能体的平均Q值



(b) SBS_2作为智能体的平均Q值

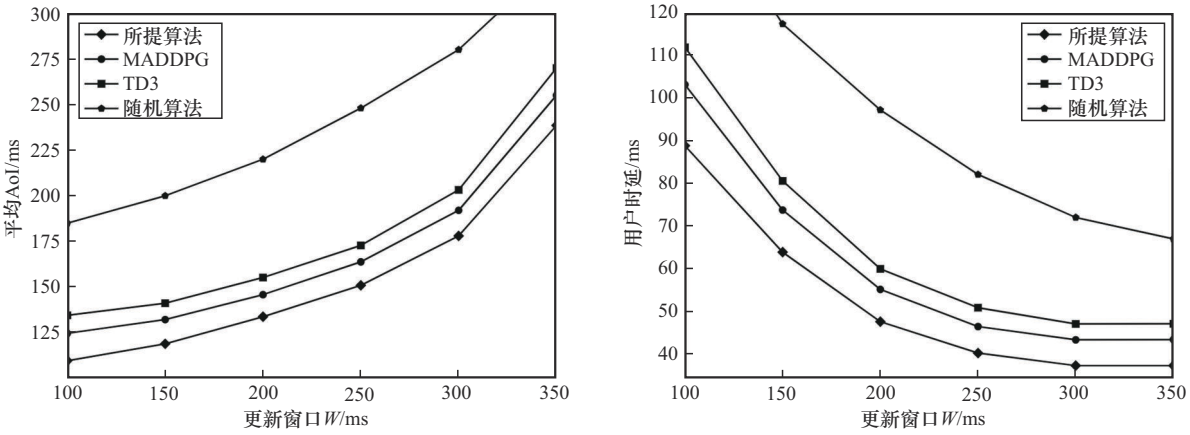


(c) MBS作为智能体的平均Q值

图5　不同智能体收敛的平均 Q 值

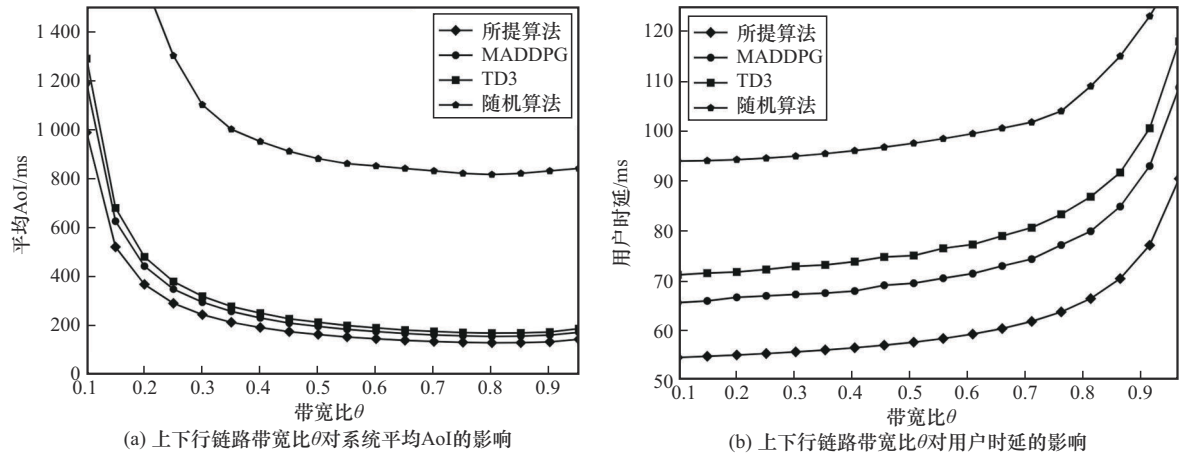
图6为SBS缓存刷新窗口阈值W对系统平均AoI和用户时延的影响。从图6(a)中可以看出，随着刷新窗口阈值的增大，位于SBS处的缓存内容的AoI也线性增大。从图6(b)则可以看出，随着缓存刷新窗口阈值的增大，SBS将缓存内容分发给车辆所产生的用户时延呈下降趋势并最后不再变化。这是因为缓存刷新窗口阈值越大，需要SBS通过上行链路与MBS进行下一次缓存内容更新的时间间隔就增大，而位于SBS缓存内容的AoI随时间线性增加，因此导致平均AoI增加。另外，虽然用户时延随着刷新窗口阈值的增加而减小，但整体系统的平均AoI依然增加，说明刷新窗口阈值的增加对平均AoI的影响大于用户时延的影响。在图6(a)中，随机算法在平均AoI上的表现明显不如其他强化学习算法。因为该算法不考虑实际网络状态，只是随机选择动作，所以在复杂网络环境下，无法提供有效的资源优化策略。本文提出的内容新鲜度保障的多智能体强化学习算法与MADDPG算法、TD3算法具有相同的上升趋势，但本文所提算法下的平均AoI与它们相比较而言分别减少了11.7%和20.6%左右。在图6(b)中，对于用户时延方面，所有强化学习算法的表现均显著优于随机算法，同时均呈现下降的趋势，但本文所提算法与另外2种算法相比在用户时延上有更好的性能。

图7为上下行链路带宽比θ对系统平均AoI和用户时延的影响。由图7(a)可知，随着带宽θ的增大，系统平均AoI整体呈现下降的趋势，在带宽比θ=0.85时系统的平均AoI最小，在之后系统平均AoI有微小的上升。与此相反的是，图7(b)中用户时延随着带宽比θ的增加而增加，在带宽比θ=0.85之后用户时延上升的趋势增大。这是因为带宽比θ决定着上下行链路频谱资源的大小，θ越大，可用的上行链路资源越多，SBS处缓存内容的AoI越



(a) SBS缓存刷新窗口阈值W对系统平均AoI的影响　　(b) SBS缓存刷新窗口阈值W对用户时延的影响

图6　SBS缓存刷新窗口阈值W对系统平均AoI和用户时延的影响



(a) 上下行链路带宽比θ对系统平均AoI的影响　　(b) 上下行链路带宽比θ对用户时延的影响

图7　上下行链路带宽比θ对系统平均AoI和用户时延的影响

小，但下行链路中的用户时延性能则相反，θ=0.85 后用户时延的增加超过了系统平均 AoI 的增长，所以图 7(a)中在 θ=0.85 之后出现略微的上升。另外，随机算法的系统平均 AoI 较高，主要原因在于其缺乏有效的决策能力。由于该算法在资源分配过程中采取随机选择，无法根据网络状态进行优化调整，从而显著增加了系统的平均 AoI。相比之下，本文所提算法能更好地从经验中学习到更好的参数，在带宽比 θ 对系统平均 AoI 的影响性能上比 MADDPG 和 TD3 算法分别提高了约 5.61% 和 5.83%。而在带宽比 θ 影响用户时延性能上，随机算法在用户时延上的表现明显不如其他强化学习算法。随机算法不具备智能决策能力，导致其用户时延较高。相比之下，本文所提算法比另外 2 种强化学习算法有更好的性能，说明所提算法能为用户提供更好的内容分发策略。

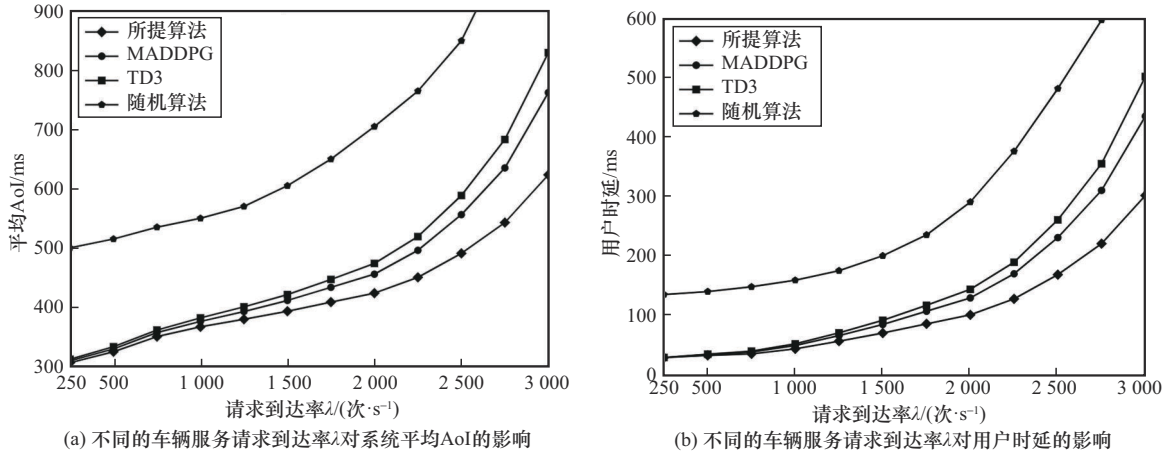
图 8 为不同的车辆服务请求到达率 λ 对系统平均 AoI 和用户时延的影响。由图 8 可知，随着车辆服务请求到达率 λ 的增加，车辆用户所获得缓存内容的平均 AoI 和用户时延都呈上升趋势。平均 AoI 的变化趋势与用户时延变化趋势相同，其原因是用户时延是主要的变化，SBS 处的缓存内容平均 AoI 的变化不明显。在 λ=2 000 后，系统的负载超出了一定的范围后，用户时延对 λ 的敏感程度更大，更容易造成更大的用户时延，同时车辆用户所获取内容的平均 AoI 也随之增大。结果表明，在所提算法下 MBS 和 SBS 能更好地为车辆用户做出缓存内容分发策略以及频谱资源分配，减少用户获得缓存内容的平均 AoI。在 λ=2 750 时，相较于其他 2 种强化学习算法，本文所提算法使车辆用户获取内容的平均 AoI 分别减少了 15.45% 和 11.47%，并且显著优于随机算法。

4.2　复杂度分析

本文提出的内容新鲜度保障的 MARL 算法通过引入注意力机制提升了智能体之间的协作能力，从而优化了内容分发决策和资源分配。为了评估算法的适用性，本文对其时间复杂度和空间复杂度进行了详细分析。

时间复杂度方面，内容新鲜度保障的 MARL 算法的主要计算来源包括注意力机制、策略更新和网络训练。其中，注意力机制用于评估相邻智能体的状态和动作对当前智能体的影响，其计算复杂度为 $O(Nd^2)$，其中 N 为智能体数量，d 为嵌入向量维度。策略更新通过采样动作空间完成，复杂度为 $O(K|A|)$，其中 K 为采样次数，$|A|$ 为动作空间的大小。此外，actor 和 critic 网络的训练复杂度为 $O(Lh^2)$，其中 L 为网络层数，h 为每层神经元的数量。综合来看，单次迭代的时间复杂度为 $O(Nd^2 + K|A| + Lh^2)$。通过引入注意力机制，内容新鲜度保障的 MARL 算法仅关注相邻智能体，避免了全局计算的高昂代价，从而保证其能够在大规模网络中高效运行。

空间复杂度方面，内容新鲜度保障的 MARL 算法的主要存储需求包括网络参数、经验回放池以及注意力计算缓存。每个智能体的 actor 和 critic 网络参数存储复杂度为 $O(Lh^2)$，经验回放池用于存储观测、动作、奖励和状态转移记录，复杂度为 $O(MS)$，其中 M 为回放池容量，S 为单条记录的存储大小。此外，注意力机制引入的临时缓存复杂度



(a) 不同的车辆服务请求到达率λ对系统平均AoI的影响　　(b) 不同的车辆服务请求到达率λ对用户时延的影响

图 8　不同的车辆服务请求到达率 λ 对系统平均 AoI 和用户时延的影响

为$O(Nd)$。总体而言，算法的空间复杂度为$O(N(Lh^2+d)+MS)$。该复杂度随智能体数量线性增长，能够满足大规模网络的存储需求。

综上，内容新鲜度保障的MARL算法在时间复杂度和空间复杂度上均保持在合理范围内。注意力机制的引入在提升协作能力的同时，有效控制了计算开销，使算法能够在动态环境下高效地完成缓存分发和资源分配任务。仿真结果进一步验证了该算法在大规模网络中的运行效率和适用性。

4.3 讨论

博弈论方法适用于多智能体系统中智能体间的协作与竞争关系。堆叠博弈可用于优化车联网中内容分发的主从决策过程，帮助SBS和车辆在内容请求与资源分配上达成动态平衡；协作博弈模型则通过激励机制进一步优化智能体之间的协作效率，从而在动态内容缓存中展现出独特优势。此外，层次化强化学习（HRL, hierarchical reinforcement learning）能够将复杂任务分解为多个子任务进行优化，在处理内容分发和缓存更新这种多阶段决策场景中表现良好，通过高层策略指导智能体选择缓存分发方式，底层策略优化资源分配，从而提升效率和鲁棒性。本文已使用注意力机制优化了多智能体间的通信效率，未来可以进一步探索基于图注意力网络（GAN, graph attention network）的协作方法。GAN在复杂网络场景中通过动态调整邻近智能体之间的影响权重，有望进一步提升内容分发和资源分配决策的全局性和准确性，为多智能体系统的优化提供更加精细化的支持。

5 结束语

为了满足具有动态内容的车辆安全应用的服务要求，本文提出了一种内容新鲜度保障的MARL算法来为车辆请求服务进行内容分发决策和资源分配决策。首先以最小化用户时延和AoI为目标将决策问题建立为一个优化问题，然后提出了内容新鲜度保障的MARL算法，将SBS当作智能体，通过智能体之间的通信来交互信息，利用注意力机制来评估相邻智能体的影响，进而优化决策。仿真表明，所提算法能有效地降低车辆用户的时延和AoI，保障车辆用户获得高新鲜度的内容，更好地适应环境的变化。未来的研究将重点关注满足车联网多种应用服务质量保障的边缘缓存，同时探索更为高效的MARL算法，以提高其在更大规模网络中的适应性和计算效率。

参考文献：

[1] NAN Z J, JIA Y J, REN Z, et al. Delay-aware content delivery with deep reinforcement learning in Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 8918-8929.

[2] ZHANG K, LENG S P, HE Y J, et al. Mobile edge computing and networking for green and low-latency Internet of things[J]. IEEE Communications Magazine, 2018, 56(5): 39-45.

[3] LI Y, XIA S C, ZHENG M Y, et al. Lyapunov optimization-based trade-off policy for mobile cloud offloading in heterogeneous wireless networks[J]. IEEE Transactions on Cloud Computing, 2022, 10(1): 491-505.

[4] LI Y, LIU J, CAO B, et al. Joint optimization of radio and virtual machine resources with uncertain user demands in mobile cloud computing[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2427-2438.

[5] JIANG K, SUN C, ZHOU H, et al. Intelligence-empowered mobile edge computing: framework, issues, implementation, and outlook[J]. IEEE Network, 2021, 35(5): 74-82.

[6] NING Z L, ZHANG K Y, WANG X J, et al. Intelligent edge computing in Internet of vehicles: a joint computation offloading and caching solution[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(4): 2212-2225.

[7] ZHOU J W, CHEN F F, HE Q, et al. Data caching optimization with fairness in mobile edge computing[J]. IEEE Transactions on Services Computing, 2023, 16(3): 1750-1762.

[8] WANG R Y, KAN Z W, CUI Y P, et al. Cooperative caching strategy with content request prediction in Internet of vehicles[J]. IEEE Internet of Things Journal, 2021, 8(11): 8964-8975.

[9] WU H H, FAN Y Z, JIN J C, et al. Social-aware decentralized cooperative caching for Internet of vehicles[J]. IEEE Internet of Things Journal, 2023, 10(16): 14834-14845.

[10] QIAO G H, LENG S P, MAHARJAN S, et al. Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks[J]. IEEE Internet of Things Journal, 2020, 7(1): 247-257.

[11] TRAN T X, POMPILI D. Adaptive bitrate video caching and processing in mobile-edge computing networks[J]. IEEE Transactions on Mobile Computing, 2019, 18(9): 1965-1978.

[12] ZHANG T K, FANG X Y, LIU Y W, et al. D2D-enabled mobile user edge caching: a multi-winner auction approach[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12314-12328.

[13] HUANG X G, XU K, CHEN Q B, et al. Delay-aware caching in Internet-of-vehicles networks[J]. IEEE Internet of Things Journal, 2021, 8(13): 10911-10921.

[14] YANG P, LYU F, WU W, et al. Edge coordinated query configuration for low-latency and accurate video analytics[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4855-4864.

[15] ZHANG M, ARAFA A, HUANG J W, et al. Pricing fresh data[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(5): 1211-1225.

[16] TIAN H, XU X L, QI L Y, et al. CoPace: edge computation offloading and caching for self-driving with deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2021, 70(12): 13281-13293.

[17] YATES R D, SUN Y, BROWN D R, et al. Age of information: an introduction and survey[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(5): 1183-1210.

[18] ZHANG S, WANG L D, LUO H B, et al. Age of information and delay tradeoff with freshness-aware mobile edge cache update[C]//Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.

[19] AHANI G, YUAN D. Optimal content caching and recommendation with age of information[J]. IEEE Transactions on Mobile Computing, 2024, 23(1): 689-704.

[20] ZHANG S, LI J J, LUO H B, et al. Towards fresh and low-latency content delivery in vehicular networks: an edge caching aspect[C]//Proceedings of the 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway: IEEE Press, 2018: 1-6.

[21] ZHANG S, LI J J, LUO H B, et al. Low-latency and fresh content provision in information-centric vehicular networks[J]. IEEE Transactions on Mobile Computing, 2022, 21(5): 1723-1738.

[22] AHANI G, YUAN D. Accounting for information freshness in scheduling of content caching[C]//Proceedings of the ICC 2020 - 2020 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2020: 1-6.

[23] 王红艳, 孙其博, 马骁, 等. 边缘辅助实时应用中信息年龄感知的任务调度[J]. 通信学报, 2024, 45(6): 144-159.
WANG H Y, SUN Q B, MA X, et al. AoI-aware task scheduling in edge-assisted real-time applications[J]. Journal on Communications, 2024, 45(6): 144-159.

[24] 张政, 谢鑫, 柏桐, 等. 信息年龄和能耗联合优化的无线体域网卸载和调度策略[J]. 通信学报, 2024, 45(9): 92-100.
ZHANG Z, XIE X, BAI T, et al. Joint optimization strategy of age of information and energy consumption for offloading and scheduling in WBAN[J]. Journal on Communications, 2024, 45(9): 92-100.

[25] ZHANG D G, WANG W J, ZHANG J, et al. Novel edge caching approach based on multi-agent deep reinforcement learning for Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(8): 8324-8338.

[26] MA M Y, WONG V W S. Age of information driven cache content update scheduling for dynamic contents in heterogeneous networks[J]. IEEE Transactions on Wireless Communications, 2020, 19(12): 8427-8441.

[27] XU C, WANG X J, YANG H H, et al. AoI and energy consumption oriented dynamic status updating in caching enabled IoT networks[C]//Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2020: 710-715.

[28] WU X W, LI X H, LI J, et al. Caching transient content for IoT sensing: multi-agent soft actor-critic[J]. IEEE Transactions on Communications, 2021, 69(9): 5886-5901.

[29] LIANG L, YE H, LI G Y. Spectrum sharing in vehicular networks based on multi-agent reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(10): 2282-2292.

[30] LIU Y, YU H M, XIE S L, et al. Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(11): 11158-11168.

[31] ZHANG S, WANG L D, LUO H B, et al. AoI-delay tradeoff in mobile edge caching with freshness-aware content refreshing[J]. IEEE Transactions on Wireless Communications, 2021, 20(8): 5329-5342.

[32] CHU T S, WANG J, CODECÀ L, et al. Multi-agent deep reinforcement learning for large-scale traffic signal control[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(3): 1086-1095.

[33] IQBAL S, SHA F. Actor-attention-critic for multi-agent reinforcement learning[C]//Proceedings of the 36th International Conference on Machine Learning. Saarland: DBLP, 2020: 2961-2970.

[34] ZHAO Y W, LI R B, WANG C Y, et al. Neighboring-aware caching in heterogeneous edge networks by actor-attention-critic learning[C]//Proceedings of the ICC 2021-IEEE International Conference on Communications. Piscataway: IEEE Press, 2021: 1-6.

[35] CHEN S Z, HU J L, SHI Y, et al. Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G[J]. IEEE Communications Standards Magazine, 2017, 1(2): 70-76.

[36] HE P, CAO L, CUI Y P, et al. Multi-agent caching strategy for spatial-temporal popularity in IoV[J]. IEEE Transactions on Vehicular Technology, 2023, 72(10): 13536-13546.

[37] ZHI Y, TIAN J, LIU Q Q, et al. Multi-agent reinforcement learning for cooperative edge caching in heterogeneous networks[C]//Proceedings

of the 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway: IEEE Press, 2021: 1-6.

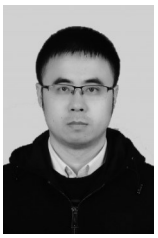
[38] ZHOU X H, BILAL M, DOU R H, et al. Edge computation offloading with content caching in 6G-enabled IoV[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(3): 2733-2747.

[作者简介]

崔亚平（1986– ），男，河南新乡人，博士，重庆邮电大学副教授，主要研究方向为智能反射面、移动边缘计算和车联网的网络功能虚拟化等。

石宏吉（1998– ），男，四川邻水人，重庆邮电大学硕士生，主要研究方向为车载网络和边缘缓存等。

吴大鹏（1979– ），男，黑龙江大庆人，博士，重庆邮电大学教授、博士生导师，主要研究方向为普适网络、IP QoS 架构、网络可靠性以及通信系统的性能评估等。



何鹏（1990– ），男，重庆人，博士，重庆邮电大学副教授，主要研究方向为移动边缘计算、分子通信。

王汝言（1969– ），男，湖北浠水人，博士，重庆邮电大学教授、博士生导师，主要研究方向为泛在网络、多媒体信息处理等。