

车载通信中基于 Q 学习的信道接入技术研究

杜艾芊^{1,2}, 赵海涛^{1,2}, 刘南杰^{1,2}

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003;
2. 南京邮电大学 网络基因工程研究所, 江苏 南京 210003)

摘要:针对基于 IEEE 802.11p 协议的车载网络 MAC 层 DCF(分布式协调功能)信道接入方法存在数据包接收率低、时延高、可扩展性差等问题,提出一种基于 Q 学习的 CW 动态调整算法-QL-CW_{min}算法。区别于现有的 BEB 算法,通过利用 Q 学习,网络节点(Agent)能够不断地与周围环境进行交互学习,根据学习结果动态地调整竞争窗口(CW),使节点总能以最佳的 CW(从周围环境中获得奖赏值最大时所选的 CW 大小)接入信道,以减少数据帧碰撞、降低端到端传输时延。仿真结果表明,采用 QL-CW_{min}算法的通信节点能快速适应车联网的未知环境,数据包接收率和数据包传输时延得到了有效改善,同时该算法能为节点接入信道提供更高的公平性,适用于各种不同负载程度的网络环境。

关键词:车载网络; BEB 算法; 竞争窗口; Q 学习算法; 分布式协调功能

中图分类号: TP301

文献标识码: A

文章编号: 1673-629X(2017)03-0085-06

doi: 10.3969/j.issn.1673-629X.2017.03.018

Research on Technology of Channel Access Based on Q -Learning Algorithm for Vehicular Communication

DU Ai-qian^{1,2}, ZHAO Hai-tao^{1,2}, LIU Nan-jie^{1,2}

(1. College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. Institute of Network DNA Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: A Q -Learning based back-off algorithm is proposed because the traditional DCF approach used for IEEE 802.11p MAC protocol to access the channel has some problems of the low packet delivery rate, high delay and the poor scalability in VANETs. The proposed algorithm, which is quite different from the traditional BEB algorithm, is adopted by the nodes (Agents) to interact with surroundings continuously and learn from each other. The vehicle nodes adjust the size of CW (Contention Window) dynamically according to the results learned from the surroundings so that the nodes can access the channel with the optimal CW eventually minimizing the packet collisions and end-to-end delay. The simulation results show that the communication nodes using the proposed algorithm can adapt to the unknown vehicular environment rapidly, and simultaneously the high packet delivery ratio, low end-to-end delay and high fairness can be achieved for vehicular network with various level load.

Key words: vehicular network; BEB algorithm; contention window; Q -Learning algorithm; DCF

0 引言

近年来,随着交通运输行业的迅速发展,汽车数量急剧增加。汽车为人们日常出行带来了便利,但也出现了安全和交通拥堵等各种问题。20世纪80年代,美国加利福尼亚大学首次提出了智能交通系统(ITS)

的概念,用以提高交通运输效率、缓解交通拥塞、减少交通事故。在智能交通系统和无线通信技术高速发展的今天,车联网应运而生,它是继互联网、物联网之后的另一个未来智慧城市的标志。车联网中,道路车辆和路边基础设施都安装有短程无线收发器,具有无线

收稿日期: 2016-03-29

修回日期: 2016-08-02

网络出版时间: 2017-01-10

基金项目: 国家“973”重点基础研究发展计划项目(2013CB329005); 国家自然科学基金资助项目(61302100, 61101105, 61201162); 江苏省基础研究计划-重点研究专项基金(BK2011027, BK2012434); 江苏省高校自然科学基金(12KJB510022, 12KJB510020)

作者简介: 杜艾芊(1991-),女,硕士研究生,研究方向为移动通信与无线技术、车联网车载通信;赵海涛,博士,副教授,研究方向为无线多媒体通信;刘南杰,博士,教授,研究方向为泛在通信、车联网、智能交通。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170110.1019.044.html>

通信功能,所以可形成一个无线网络,即车载自组织网(VANET)。VANET 是移动自组织网的子类,没有固定的拓扑结构,车辆可通过 V2V(车与车)通信或 V2I(车与路边基础设施)通信获取信息和服务。VANET 通过车-车通信和车-路通信实现人-车-路的协同,有效改善了交通安全,提高了交通效率,为用户提供娱乐和 Internet 接入服务等。

IEEE802.11p 又称 WAVE(Wireless Access in the Vehicular Environment),主要用于车载通信,是由 IEEE802.11 标准扩充的通信协议。IEEE802.11p 针对车载环境对 IEEE802.11 的物理层和 MAC 层的相关参数做了些许调整,因而能更适用于车载环境中的无线通信。IEEE802.11p 是 WAVE 协议栈的底层协议,已广泛应用于 V2V 通信。在任一网络环境中,通信协议栈的重要因素之一就是 MAC 层,而 IEEE802.11p MAC 层主要解决的是车辆对信道接入的竞争问题,它决定了某一时刻允许哪一节点接入无线信道。由于节点的高速移动性、通信环境的快速变化性及节点密度和节点分布的多变性等,对 VANETs 共享无线信道的接入控制提出了挑战。因此,设计高可靠性的 MAC 协议对 VANETs 尤为重要。为 VANET 环境设计 MAC 协议所面临的挑战主要有:在车辆位置和信道特征不断变化的 VANET 中,实现既高效又公平的信道接入;对不同密度的交通流具有可扩展性;能满足各种不同的应用需求。

1 相关文献

文献[1]提出一种基于邻居节点数估计的最小竞争窗口调整算法-Adaptive CW_{min}。算法改变了 CW 的调整规则,并根据网络信道的使用情况动态地调整 CW_{min}。通过估计车载网中的竞争节点数来动态地选择合适的 CW_{min},若数据传输成功,则根据竞争节点数确定 CW_{min};若失败,则通过估计车辆密度来控制竞争窗口的增加。同时还推导出最大退避阶数、信道由于碰撞被检测为繁忙的平均时间和竞争节点数这三个参数与最优 CW_{min} 的函数关系,节点成功发送数据后,根据函数计算出适应车载网络状况的最优 CW_{min} 值。利用文中提出的算法在数据包重传之后选择合理的 CW,缩短了竞争节点等待重传的时间,增加了网络吞吐量。

文献[2]提出了基于统计次数的退避算法-new-BEB 和基于相对距离的退避算法-RBA。newBEB 算法中设定了一个门限值,即发送节点传输成功和传输失败的最大次数。当节点连续发送成功的次数超过传输成功的最大次数时,就增加竞争窗口值,降低其竞争信道的能力,而当节点连续发送失败的次数超过传输

失败的最大次数时,就减少竞争窗口值,增强其竞争信道的能力。通过仿真对比分析,newBEB 算法有效提高了节点接入信道的公平性。RBA 算法中,每个节点根据自己与邻居节点距离的平均值动态地调整竞争窗口的大小,仿真结果表明 RBA 算法提高了节点接入信道的公平性,降低了丢包率,在一定程度上提高了网络吞吐量。

文献[3]提出一种 CW 的控制方法-DBM-ACW 方法(基于密度调整 CW 的方法)。根据信道拥塞的严重程度,选择不同的 CW 值倍乘系数,或重设为 CW_{min}。信道十分拥塞时,CW 值的倍乘系数选择上限值,可减少节点选择相同退避数的概率;当信道密度降低时,CW 值的倍乘系数选择下限值或重设为 CW_{min},避免节点在信道占用率较低时等待较长的时间接入信道。经仿真对比分析,文中提出方法在网络密度较大时,性能优势尤为突出。

文献[4]提出一种基于距离动态调整 CW 值的方法,适用于在网络负载较重的车载自组织网中广播实时性紧急消息。文中推导出某节点和前一节点之间的距离 d 和动态竞争窗口 CW_{*d*} 之间的关系,利用这一关系为不断移动的车辆节点动态地分配不同的 CW 值,可减少由于碰撞需要重传数据包的次数。此外,还能降低数据包碰撞概率、端到端时延及网络负载等,最终使带宽得到有效利用。仿真结果表明,此方法在高速公路交通流中就吞吐量、端到端时延和网络负载而言,网络性能得到有效改善。

传统的退避算法虽然解决了信道竞争的问题,但同时也存在一定的缺陷。节点每成功发送一次数据,就把当前的 CW 降到最小值,使节点误以为一次发送成功就表示当前信道竞争情况不激烈,同样,一次发送失败就认为当前信道竞争激烈,碰撞增加,相应的竞争窗口成倍增加^[5]。这种方式并没有真正反映信道上的竞争情况。尤其是节点个数较多网络负载变严重时,成功发送完数据的节点把 CW 调为最小值,多个节点又有相同的 CW 值(CW_{min}),会引起更多的碰撞。且碰撞和退避会浪费时间,严重影响网络的整体吞吐量。另外,数据发送成功的节点立刻将 CW 值置为最小,而数据发送失败的节点的 CW 值成倍增加。之后的一段时间内,CW 值小的节点再次成功接入信道的概率增加,而 CW 值大的节点 CW 值会继续增加,若 CW 值小的节点连续发送数据的话,就会一直占用信道,而其他节点的数据就无法发送出去,最终造成严重的信道接入不公平现象。除了接入信道不公平的问题,在时延方面也存在一定的缺陷。IEEE 802.11p 在 MAC 层中规定,初始竞争窗口值为 15(aCW_{min})^[6]。但是,数据流量负载过高(网络密度较高)时,成功发送数据帧

后恢复为初始竞争窗口会使碰撞率增加,若节点在预定义时间段内无法成功发送数据帧对应的ACK消息,发送节点就会增加竞争窗口,重传数据帧,这样会使时延增加。即使在初次尝试发送数据时设定最佳的竞争窗口可避免增加时延,但节点成功发送数据后,竞争窗口又恢复为15,尤其是在网络负载过高时,极其不宜采用这种方法。

2 基于Q学习的CW动态调整算法

针对上述问题,文中在传统信道接入机制的基础上,引入强化学习,基于强化学习中的Q-Learning算法提出了新的信道接入方法-QL-CW_{min}方法。

2.1 Q学习的基本原理

强化学习是机器学习中尤为重要的一种,它是智能系统从环境状态到行为映射的学习过程,是解决智能系统寻优问题的有效工具。能够感知环境的Agent不断在环境中学习,根据环境给予的反馈信号,总选择执行能达到目标的最优动作。在强化学习中,Agent在环境中的学习过程是一种试探评价过程,它的基本原理是^[7]:Agent在环境中学习时,选择一个动作作用于环境,环境状态受该动作影响而发生变化,同时会产生一个强化信号(奖或惩),并将此信号反馈给Agent,Agent会根据强化信号和环境的当前状态再选择执行下一个动作,如果Agent的某一动作策略会使Agent从环境中获得正的奖赏(强化信号),那么Agent此后选择执行这个动作策略的趋势就会加强,它的最终目的是Agent总选择执行能从环境中得到最大累积奖赏值的动作。

Q-Learning算法是强化学习算法中最典型的一种,也是目前应用最为广泛的一种。Q-Learning算法在环境条件未知的情况下最为有效,对环境的先验知识要求不高,它使Agent在马尔可夫决策过程中具备利用经历过的动作序列总选择最优动作的能力。它不需要环境模型,Agent在动态环境中通过交互试错不断调整行为。Agent不断探索环境,在每一环境状态和可能的动作之间建立一个Q值列表(Q表),它学习的是每个状态-动作对的评价值-Q值($Q(s_t, a_t)$), $Q(s_t, a_t)$ 是Agent在状态 s_t 下根据策略选择执行动作 a_t ,并循环执行所得到的累积奖赏值。Q-Learning算法的最优策略是使 $Q(s_t, a_t)$ 的累积奖赏值最大化,所以Q学习的最优策略表达式为^[8]:

$$\pi^*(s_t) = \arg \max_{a_t} Q(s_t, a_t) \quad (1)$$

即Agent只需考虑当前状态和当前可选的动作,按照策略选择执行使 $Q(s_t, a_t)$ 最大化的动作。

文中所提基于信道接入方法-QL-CW_{min},通过动

态调整竞争窗口来解决碰撞率和时延的问题,利用Q-Learning算法学习最佳的竞争窗口。由于邻近节点之间互换信标消息可获得邻居节点的位置信息,所以假设每个节点已知其一跳邻居节点的位置信息,在节点成功发送数据帧后,环境给予节点一个正的奖赏,若发送失败,则给予负的奖赏。在网络负载较低时,使节点利用学习所得的最佳CW选择以较小的CW接入信道避免增加时延;网络负载较高时,则利用较大的CW接入信道防止碰撞。QL-CW_{min}算法可动态地调整竞争窗口,能以较低的时延发送数据,提高了数据包接收率和竞争效率,减少了信道接入时延。

2.2 QL-CW_{min}算法的状态-动作对映射

整个车载自组织网络即Agent学习的环境,网络中的每个车辆节点即Agent,车辆节点在网络中接入信道时所采用的竞争窗口即Agent学习环境的环境状态,由此车辆节点可能采用的所有竞争窗口集即Agent学习环境的状态空间。由于节点在网络中接入信道的竞争窗口通常为2的指数幂减1,因此竞争窗口集为 $\{15, 31, 63, 127, 255, 511, 1023\}$,竞争窗口初始值CW_{min}为15,最大值CW_{max}为1023^[9]。每一Agent可执行的动作有:增加(I)、保持(K)、减少(R)。“增加”即增大竞争窗口,“保持”和“减少”则分别是保持竞争窗口大小不变和减小竞争窗口。节点每执行一个动作后,环境状态就发生状态转移。在网络环境中不断探索学习的过程中,每一节点在状态-动作对之间都维护一个Q表,Q表中包含Q值 $Q(s_t, a_t)$,Q值的变化范围为-1到1。其中, s_t 为当前竞争窗口的大小, a_t 为节点可能执行的动作。每发送完一个MAC帧后,节点根据发送状态从网络环境中获得一个奖赏值。若发送成功,节点得到一个正的奖赏,若发送失败(本协议中定义MAC层重传次数不超过4,即数据重传4次后,发送节点还是接收不到数据帧对应的ACK消息,则定义此次发送失败),节点则得到一个负的奖赏。丢包主要是由于其他数据包发生碰撞造成的,通过对奖赏值进行评估,节点自适应地调整其竞争窗口大小,总选择执行能使累积奖赏值Q值最大化的最优动作。

2.3 QL-CW_{min}算法Q值函数更新

在Agent与环境不断交互学习的过程中,节点接入信道可能执行的动作有:增加(I)、保持(K)、减少(R)。状态空间为 $\{15, 31, 63, 127, 255, 511, 1023\}$ 。当竞争窗口为最小值时,竞争窗口无法继续减少;同样地,当竞争窗口为最大值时,竞争窗口无法继续增加^[10]。图1为节点在网络环境中学习的状态转移图。

VANETs中,节点采用QL-CW_{min}算法发送MAC数据帧的过程中,利用状态-动作对的值函数 $Q(s_t, a_t)$ 进行迭代,并利用奖赏作为估计函数来选择下一动

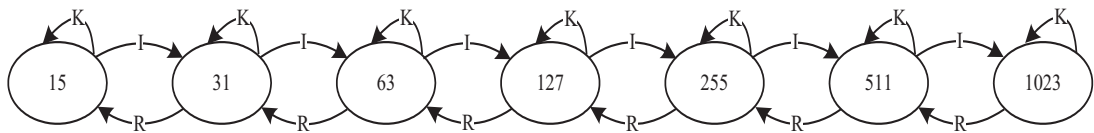


图 1 状态转移图

作,对 Q 函数进行优化,通过多步迭代学习逼近最优值函数。节点每发送一次数据帧,就更新一次 Q 表,更新 Q 值的表达式即 Q 学习的迭代公式为:

$$Q(s_t, a_t) \leftarrow \alpha \times \{R + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})\} + (1 - \alpha) \times Q(s_t, a_t) \tag{2}$$

其中, α 为学习率,是 Agent 在环境中的学习步长,用于控制学习速度。 α 值越大, Q 值收敛越快。由于 MAC 数据帧发送较为频繁,0.6 足以反映网络拓扑的变化程度,所以文中取 α 为 0.6。 γ 为折扣因子, $\gamma \in [0, 1]$,体现了 Agent 对以后环境所给予奖励的重视程度,取值越大表示越重视以后的奖励,反之,则只在乎眼前的奖励。文中取 γ 为 0.9。

车辆节点在 VANETs 中初次接入信道发送数据时,会首先初始化 $Q(s_t, a_t)$ 的值,然后根据探索策略在状态 s_t 时选择执行动作 a_t ,得到下一状态 s_{t+1} 及其奖赏值 R ,之后根据奖赏值通过式(2)更新 Q 值,一直循环执行直到实现目标状态或达到限制的迭代次数。其中奖赏值 R 计算如下:

$$R = \begin{cases} R_{CW}, & \text{发送成功} \\ -1, & \text{发送失败} \\ 0, & \text{当前状态正在发送数据} \end{cases} \tag{3}$$

其中, R_{CW} 表示选择当前的 CW 值接入信道成功发送数据所获得的正奖赏。发送失败,奖赏值为-1,若当前状态正在发送数据,奖赏值为0。

表 1 中定义了选择各不同大小的 CW 值成功发送数据所获得的不同奖赏值。成功发送数据所选的 CW 值越小,得到的奖赏值就越大,而网络负载过高时,节点从环境中获得负的奖赏从而增加竞争窗口,这样能使节点充分利用信道资源。

表 1 不同 CW 对应的奖赏值

CW 大小	R_{CW}
15	1
31	6/7
63	5/7
127	4/7
255	3/7
511	2/7
1 023	1/7

节点每从环境中获得一次奖赏,就按照式(2)更新一次 Q 值,式中的 $\max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$ 表示执行动作 a_{t+1}

后所获得的最大 Q 值^[11],即到 s_{t+1} 状态为止节点从环境中所获得的最大累积奖赏值。 s_{t+1} 表示选取执行动作 a_{t+1} 后的状态。例如,竞争窗口大小为 15 时,节点接入信道发送数据发生碰撞,无法成功发送数据,下次再发送数据就选择执行“增加”动作,增加竞争窗口大小,此时状态转移为{31}。更新 Q 值的算法流程如图 2 所示。

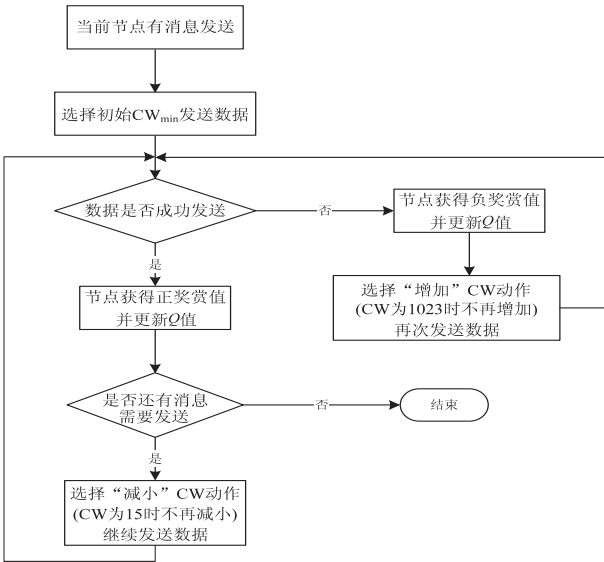


图 2 基本流程图

2.4 QL- CW_{min} 算法的收敛性

强化学习中,“探索”是指 Agent 要尽可能地经历所有的状态-动作对,从而获得全面充分的经验知识,保证学习过程能收敛到最优的 Q 值函数,但是过度“探索”会引入冗余信息,浪费存储资源和计算资源,最终影响学习速度。“利用”则是 Agent 为了从环境中获得较高的奖赏值,总是根据当前的 Q 表选择执行可以获得高奖赏值的动作,而不愿冒险去尝试可能会产生更高奖赏值但也可能产生低奖赏值的动作^[12]。所以寻求“探索”和“利用”间的平衡对保证学习过程能快速收敛到最优 Q 值函数非常重要,Agent 需要不断“探索”次优动作从而使“利用”趋向全局最优。

QL- CW_{min} 算法中,节点在网络环境中学习所用的探索策略为强化学习算法中应用较为广泛 ϵ -greedy 动作选取机制,每个 Agent 节点要执行的第一个动作是将其 CW 值初始化为 15,当 Agent 对自己所处的网络环境一无所知时,采用最小的 CW 值是最佳选择。此后节点以概率 ϵ 进行探索,寻求新的可能会产生更高奖赏值但也可能产生低奖赏值的动作,以概率 $1 - \epsilon$

选择当前 Q 值最高的动作(利用)。由于节点接入信道并成功发送数据所选用的 CW 越小,Agent 得到的奖赏就越多,只要当前所选的 CW 能成功发送数据,节点就绝不会再增加 CW 。当 CW 大于 15,而网络负载降低时,QL- CW_{min} 算法也会通过探索将 CW 重设为 15,即 QL- CW_{min} 算法总能使节点在网络环境中通过“探索”和“利用”将 CW 调整为最佳值。

收敛问题也是强化学习算法所研究的一个重要问题^[13],Watkins 与 Dayan 利用随机过程和不动点理论给出:

- (1)学习过程具有 Markov 性;
- (2)所有的状态-动作对被无限次访问;
- (3) Q 表中能存储所有状态 - 动作对的 Q 值函数,每个元素分别对应于一个状态 - 动作对;
- (4)学习率 α 满足一定的取值条件: $0 \leq \alpha_t \leq 1$,

$$\sum_{t=0}^{\infty} \alpha_t = \infty, \sum_{t=0}^{\infty} \alpha_t^2 < \infty。$$

以上四个条件都满足时, Q 学习过程可收敛到最优状态-动作对值函数 Q^* 。由此可见,QL- CW_{min} 满足收敛的所有条件。

3 仿真结果

文中对高速公路场景进行仿真,仿真环境如表 2 所示。

表 2 仿真参数

参数	值
高速公路	2 000 m,4 车道(每方向两车道)
节点数	10 ~ 100
移动性	最大速度为 80 km/h
MAC	IEEE 802.11p(11Mbps)

文中通过仿真从 VANET 的广播帧接收率、端到端传输时延和数据帧碰撞率等方面对 $CW_{min}-15$ 算法、 $CW_{min}-31$ 算法、Adaptive CW_{min} 算法和提出的 QL- CW_{min} 算法进行了对比分析。

图 3 分别为 VANET 中采用四种退避算法的广播帧接收率随车辆密度的变化趋势。

从图中可以看出,车辆密度较小时,由于车辆节点总会以最小的竞争窗口接入信道,且数据帧发生碰撞的概率很小,所以广播帧接收率都较接近 1。但随着车辆密度的增加,广播帧接收率都呈下降趋势,而采用文中所提出的 QL- CW_{min} 算法的广播帧接收率不仅优于其他算法,且下降趋势较为平稳。因为车辆节点在 VANET 环境中不断交互试错,总选择 Q 值最高的竞争窗口即最佳窗口接入信道,所以其广播帧接收率不会随着车辆密度的增加而呈急剧下降的趋势。

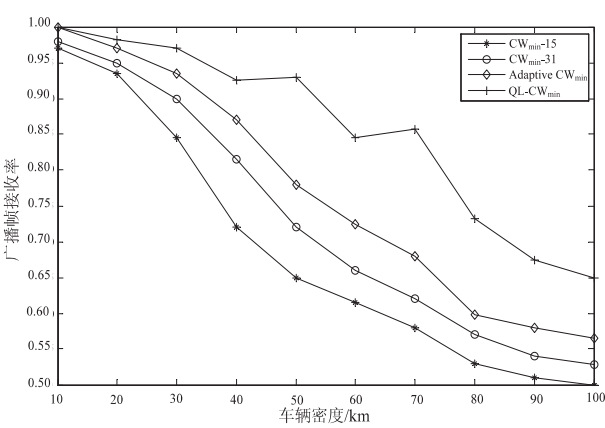


图 3 广播帧接收率与车辆密度的关系曲线

图 4 和图 5 分别是采用四种退避算法后 VANET 中的端到端传输时延和数据帧碰撞率随车辆密度的变化趋势。

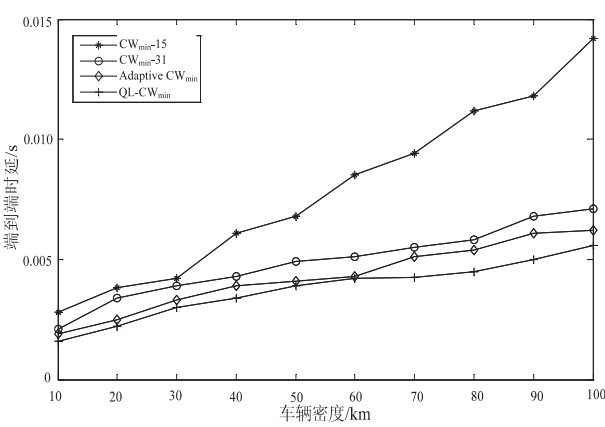


图 4 端到端时延与车辆密度的关系曲线

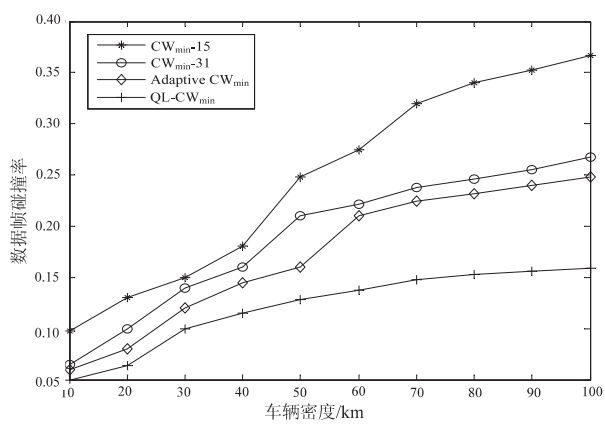


图 5 数据帧碰撞率与车辆密度的关系曲线

从图中可见,文中提出的 QL- CW_{min} 算法与其他三种退避算法相比,端到端传输时延和数据帧碰撞率都较小,因此采用 QL- CW_{min} 算法能为节点接入信道提供更高的公平性,从而能适用于各种不同负载程度的车载网络。

总的来说,在节点接入信道发送数据发生碰撞需要进行退避的过程中,车辆节点利用 Q 学习算法与周围环境不断交互,根据网络环境反馈的奖赏信号,动态地调整竞争窗口,使节点下次发送数据时总能以最佳

的 CW 值接入信道,提高了数据成功发送的概率,减少了退避次数。就数据包接收率及端到端传输时延,QL-CW_{min}算法性能与其他传统的退避算法相比都得到有效改善,尤其是 QL-CW_{min}算法完全不同于以往传统的退避算法^[14]在节点成功发送完数据后就将 CW 值恢复为 15,从不考虑网络负载情况,而是使节点能经历所有的状态-动作对,根据奖赏值总以较大的概率选择能成功发送数据且能获得较高奖赏值的 CW。因此,QL-CW_{min}算法显著提高了节点接入信道的公平性,且还能适用于不同负载程度的网络环境。

4 结束语

文中将强化学习中的 Q 学习算法引入到 VANETs MAC 层中,详细设计了车辆节点与车载网络环境交互学习时所处的状态和可执行的动作,以及状态-动作对的映射关系和动作策略,提出了 QL-CW_{min} 退避算法,并通过仿真对该算法和现有的其他退避算法进行了对比分析。QL-CW_{min}算法就广播帧接收率、端到端时延、数据帧碰撞率、节点公平性及可扩展性而言,性能得到显著改善。在 ε -greedy 动作选取机制中 ε 值固定的情况下,节点在环境中探索新动作时,选择执行各个动作的概率一样,这样选择最坏动作的概率也很大。后续研究中,为了有效地平衡“探索”和“利用”,将会尝试动态调整 ε 值,当前 Q 值最高的动作赋予最高的概率,而探索的新动作要根据其奖赏值大小赋予不同的概率,从而使研究结果更精确,考虑因素更全面。另外,已提出的基于 Q 学习的信道接入退避方法是针对单智能体的学习过程的,单智能体的学习方法存在一些问题。例如,智能体对环境仅部分感知、学习搜索空间太大、学习效率低等。因此,针对这些问题,将进一步研究基于多智能体 Q 学习的接入信道退避方法,有效改善节点接入信道的各方面性能。

参考文献:

- [1] Reinders R, Eenennaam M, Karagiannis G, et al. Contention window analysis for beaconing in VANETs[C]//2011 7th international wireless communications and mobile computing conference. [s. l.]:[s. n.],2011:1481-1487.
- [2] Alasmary W, Zhuang W. Mobility impact in IEEE 802.11p infrastructureless vehicular networks[J]. Ad Hoc Networks, 2012,10(2):222-230.
- [3] Balador A, Calafate C T, Cano J C, et al. Reducing channel contention in vehicular environments through an adaptive contention window solution[J]. Wireless Days, 2013, 30(12):1-4.
- [4] Lee Gil-Won, Ahn Seung-Pyo, Kim Dong-Seong. Contention window allocation scheme for V2V[C]//2013 international conference on advanced technologies for communications. [s. l.]:[s. n.],2013:501-505.
- [5] Booyesen M J, Zeadally S, van Rooyen G J. Survey of media access control protocols for vehicular ad hoc networks[J]. IET Communications,2011,5(11):1619-1631.
- [6] Shi C, Dai X, Liang P, et al. Adaptive access mechanism with optimal contention window based on node number estimation using multiple thresholds[J]. IEEE Transactions on Wireless Communications,2012,11(11):2046-2055.
- [7] Lampton A, Valasek J. Multiresolution state-space discretization method for Q-Learning[C]//American control conference. [s. l.]:[s. n.],2009:10-12.
- [8] 赵 昀,陈庆伟,胡维礼. 一种基于信息熵的强化学习算法[J]. 系统工程与电子技术,2010,32(5):1043-1046.
- [9] Wang Q, Leng S, Fu H, et al. An IEEE 802.11p-based multi-channel MAC scheme with channel coordination for vehicular ad hoc networks[J]. IEEE Transactions on Intelligent Transportation System,2012,13(2):449-458.
- [10] 魏李琦,肖晓强,陈颖文,等. 基于相对速度的 802.11p 车载网络自适应退避算法[J]. 计算机应用研究,2011,28(10):3878-3880.
- [11] 杜春侠,高 云,张 文. 多智能体系统中具有先验知识的 Q 学习算法[J]. 清华大学学报:自然科学版,2005,45(7):981-984.
- [12] 陈学松,杨宜民. 强化学习研究综述[J]. 计算机应用研究,2010,27(8):2834-2838.
- [13] 黄玉清,王英伦. 支持服务区分的多智能体 Q 学习 MAC 算法[J]. 计算机工程,2013,39(8):112-116.
- [14] van Eenennaam M, Remke A, Heijenk G. An analytical model for beaconing in VANETs[C]//Vehicular networking conference. [s. l.]:IEEE,2012:9-16.