

一种基于 Q-Learning 策略的自适应 移动物联网路由新算法

张德干^{1,2}, 葛 辉^{1,2}, 刘晓欢^{1,2}, 张晓丹³, 李文斌^{1,2}

(1. 天津理工大学 天津市智能计算及软件新技术重点实验室, 天津 300384;
2. 天津理工大学 计算机视觉与系统省部共建教育部重点实验室, 天津 300384; 3. 中国科学技术信息研究所, 北京 100038)

摘 要: 针对移动物(车)联网的路由问题,通过对车辆的运动特点及造成链路断裂的原因进行的详细分析,我们建立了链路维持时间模型,并将维持时间作为设计路由算法的重要参数. Q-Learning 作为一种启发式机器学习策略,能够通过与周围环境交互来动态地调整路由路径. 基于此,我们设计了一种自适应的路由新算法. 它将学习任务分散在每一个车辆节点中,通过周期性的与周围节点交换信标信息来维护可靠的路由路径. 利用 NS-2 模拟器对该算法的性能进行了评估,结果表明,在不同的网络场景中,该算法在递交率、端到端的延时以及平均跳数等方面均表现出很好的效果.

关键词: 机器学习; 移动物联网; 拓扑; 动态; 路由

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2018)10-2325-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.10.003

A Kind of New Routing Algorithm with Adaptivity for Mobile IOT Based on Q-Learning

ZHANG De-gan^{1,2}, GE Hui^{1,2}, LIU Xiao-huan^{1,2}, ZHANG Xiao-dan³, LI Wen-bin^{1,2}

(1. Tianjin Key Lab of Intelligent Computing & Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China;

2. Key Laboratory of Computer Vision and System (Tianjin University of Technology), Ministry of Education, Tianjin University of Technology, Tianjin 300384, China;

3. Institute of Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: In order to solve the routing problem of mobile IOT (IOV), based on our analyzing the details about motion characteristics of the vehicle and the reasons that cause links down, we set up link model of the duration time and using the duration time as key parameter to design the new routing method. Q-Learning as a kind of heuristic machine learning strategy is able to dynamically adjust the routing path through interaction with the surrounding environment. So a kind of new routing algorithm with adaptivity for mobile IOT based on Q-learning has been presented in this paper. It distributes the learning task into each vehicle node and maintains the reliable routing path by continuously exchanging the beacon information with the neighbor nodes. With the NS-2 simulator, the performance of the algorithm is tested. The results show that it has better performances on delivery, end-to-end delay and average hops in many mobile applications.

Key words: machine learning; mobile IOT; Topology; dynamic; routing

1 引言

随着移动物(车)联网的发展,车载娱乐以及车间通信成了汽车制造商为用户提供的必备功能^[1-4]. 在

VANETs (Vehicular Ad Hoc Networks)中,无线通信可以分为两种,一种是车辆与固定基础设施的(V2I)通信,另一种为车辆与车辆之间(V2V)通信. 由于V2V不需要基础设施且部署灵活的特性使得其受到许多研究者

收稿日期:2017-08-16;修回日期:2017-12-09;责任编辑:孙瑶

基金项目:国家自然科学基金(No. 61571328);天津市重大科技专项(No. 15ZXDSCX00050, No. 16ZXFVGX00010);天津市科技支撑重点项目(No. 17YFZCGX00360);天津市自然科学基金(No. 15JCYBJC46500);天津市科技创新和131人才团队(No. TD12-5016, 2015-23, No. TD13-5025)

的青睐. 与传统的 MANETs 有所不同的是, VANETs 有以下几个特点^[5-10]. 首先, 在 VANETs 环境下, 作为网络节点的车辆运动速度快, 这将引起网络拓扑的频繁变化, 节点间链路维持时间短暂, 使得链路极不可靠. 而且随着车辆的分布情况不同变化很大. 其次, 车辆节点的运动被限制在道路上, 受到很多因素的影响, 比如速度限制等. 第三, 车辆本身能够提供足够的电能以及强大的运算能力, 使得在 MANETs 中能量的限制不再是一种重要的影响因素. 由于 VANETs 的拓扑变化频繁以及链路不可靠的特点, 使得传统的基于 MANETs 的路由算法, 很难被应用在 VANETs 网络中. 为了克服这一问题, 近些年来, 一些基于地理位置的路由算法被提出^[11-15], 比如 TFOR 和 GyTAR 等. 这些路由算法不需要考虑拓扑的变化, 利用 GPS 来定位目的节点, 并且通过对交通流的估计来设计路由算法, 但是却并没有考虑节点间链路的可靠性. 一些可靠的路由算法被提出^[16-18], 比如 SLBF 和 EG-RAODV, 这些算法通过加入对节点间链路的可靠性以及包错误率参数来设计路由算法, 能够在一定程度上提高数据包的递交率, 但延时时间得不到保证.

考虑到车辆都运动都是按照固定的路线行驶, 两个车辆的行驶方向, 车间距离, 速度以及加速度等这些因素, 均能造成两个节点间链路的断裂. 随着近几年来对路由算法的不断研究^[19-23], Q-Learning 算法作为一种自学习的算法通过不断与外界交互信息, 其能够在动态的环境中找出一条从源节点到目的节点的最短路径. 基于这一思想, 我们通过改进 Q-Learning 算法设计出一种自适应路由算法 RSAR, 其能够在保证每一跳链路可靠性的前提下自适应地调整 Q-Table 表, 使其适应 VANETs 动态的网络拓扑.

2 链路可靠模型

2.1 系统模型

为了能够有效的评估节点间链路的质量, 考虑到高速公路一般为一条笔直的直线道路, 由于广播距离远大于道路宽度, 我们假设道路宽度对我们选择下一跳转发节点的影响很小, 因此忽略道路宽度. 我们将高速公路模型化为如图 1 所示的一种情形.

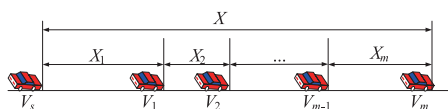


图1 道路模型

我们仍然假设在此条道路上, 车辆运动有加速、减速、变道以及超车等现象的发生. 在这条道路上, 车间距离服从于对数正态分布, 即 $X_i \in \log N(\mu_i, \delta_i)$. $X_i = \{X_i$

$(m), m=0, 1, 2, \dots\}$ 是一个服从于对数正态分布的随机变量(如图 1 所示). X_i 代表车辆 i 与 $i+1$ 之间的距离. $X_i(m)$ 是一个随机变量, 代表着节点 i 在时间 m 时的车间距离. 在图 1 中, 我们以 V_s 节点为参考节点, 那么 V_s 到任一节点的距离用 X 表示, 其中, $X = X_1 + X_2 + \dots + X_m$ 即 $X = \sum_{i=1}^m X_i$, 那么 X 也服从于对数正态分布.

引理 1 假设 $X \in \log N(\mu, \sigma)$, 那么随机变量 $T = \sqrt{aX+b} + c$ 服从于对数正态分布, 其中 $a, b, c \in R, a, b, c \neq 0$, 且 $aX+b \geq 0$.

证明 参见文献[13].

引理 2 假设 X 服从于对数正态分布 $X \in \log N(\mu, \sigma)$, 那么随机变量 $T = aX + b$ 也服从于对数正态分布. 其中, $a, b, c \in R$, 并且 $a, b, c \neq 0$.

证明 同理, 我们根据引理 1, 令 G_T 为 T 的概率分布函数, 将其带入引理 1 中证明正确性.

2.2 一跳范围内链路持续时间模型

考虑到车辆的运动总是按照固定的道路行驶, 当两个节点链路出现断开状态时, 主要由两种情况所造成: 两个车辆同向而行、两个车辆背道而行. 以车辆 i 为参考节点, 作为发送节点. 而其他情况车辆 i 与 j 均有着最长的链路通信时间. 对于两个车辆同向而行的情形, 假设在时间 $t_0 = 0$ 时, 车辆 j 处于车辆 i 的一跳通信范围内, 且车辆 j 位于车辆 i 之前, 初始车间距离 X 是一个随机变量. 车辆的最大通信半径 R 为一固定常量. 初始时刻, X 满足 $0 \leq X < R$, 根据系统模型, 在高速道路上, 车辆存在加速, 减速超车的情况发生. 道路上规定的最大限制速度设为 v_m , 道路上的所有车辆速度必须小于或者等于 v_m . 假设任一车辆初始时的车辆的加速度为 $a(0)$, 速度为 $v(0)$. 当 $t \geq 0$ 时, 定义为加速度为 $a(t)$, 速度为 $v(t)$. 在时间 t 时的加速度如下几种情况:

(1) 如果 $a(0) = 0$, 对于所有 $t \geq 0$ 时, 有

$$a(t) = 0 \quad (1)$$

(2) 如果 $a(0) > 0$, 有

$$a(t) = \begin{cases} a(0), & \text{满足 } t \leq \frac{v_m - v(0)}{a(0)} \\ 0, & \text{其他} \end{cases} \quad (2)$$

(3) 如果 $a(0) < 0$, 有

$$a(t) = \begin{cases} a(0), & \text{满足 } t \leq \frac{-v(0)}{a(0)} \\ 0, & \text{其他} \end{cases} \quad (3)$$

从以上分析我们假设, 当在 $t_0 = 0$ 时, 如果此时的加速度 $a(0)$ 也为 0, 认为其瞬时的加速度也为 0, 即 $a(t) = 0$. 而对于式(2)和式(3)来说, 只要速度没有达到最大限制速度或者是减小为 0 时, 则认为其加速度仍然为 $a(0)$, 否则加速度变为 0. 考虑到车辆的初速度为

$v(0)$,那么在时间 t 时,其瞬时速度 $v(t)$ 在时间 t 时的速度由以下公式定义

$$v(t) = v(0) + \int_0^t a(u) du \quad (4)$$

其中, $u \in [0, t]$, $a(u)$ 是在时间 u 时的加速度.

下面结合式(4),来计算瞬时速度:

(1) 如果 $a(0) = 0$,那么对于所有的 $t \geq 0$,有 $v(t) = v(0)$

(2) 如果 $a(0) > 0$,那么有

$$v(t) = \begin{cases} v(0) + a(0)t, & \text{满足 } t \leq \frac{v_m - v(0)}{a(0)} \\ v_m, & \text{其他} \end{cases} \quad (5)$$

(3) 如果 $a(0) < 0$,有

$$v(t) = \begin{cases} v(0) + a(0)t, & \text{满足 } t \leq \frac{-v(0)}{a(0)} \\ v_m, & \text{其他} \end{cases} \quad (6)$$

根据以上定义,任一车辆的速度为 $v(x)$ 在时间间隔 $[0, t]$ 内行驶的距离被定义为

$$S(t) = \int_0^t v(x) dx \quad (7)$$

根据以上定义,我们能够计算出在时间 t 时车辆 i 与 j 行驶的距离. 假设车辆 i 与车辆 j 的初始速度与加速度分别为 $a_i(0), v_i(0), a_j(0), v_j(0)$. 在时间 t 时的瞬时加速度与速度分别为 $a_i(t), v_i(t)$ 以及 $a_j(t), v_j(t)$. 根据式(7),我们就可以得到在时间间隔 $[0, t]$ 内车辆 i 与 j 的各自的行驶距离为

$$S_i(t) = \int_0^t v_i(x) dx, S_j(t) = \int_0^t v_j(x) dx$$

通过以上定义,当初开始时车辆 i 与 j 的车间距离为 X 时,那么在时间 t 时 i 与 j 的距离 $d_{i,j}$ 为

$$d_{i,j} = \begin{cases} S_j(t) + S_i(t) + X, & \text{方向相反} \\ S_j(t) - S_i(t) + X, & \text{方向相同} \end{cases} \quad (8)$$

从式(8)中可以明显的看出,当 $d_{i,j} > R$ 时,此时的链路将是断开的.

我们来分析两个车辆朝着相反的方向行驶时的链路维持时间. 当两个车辆满足条件(9)时

$$S_j(t) + S_i(t) + X = R \quad (9)$$

能够计算出最大的链路维持时间 t . 考虑, $S_j(t) + S_i(t) = \frac{1}{2} \cdot a_r t^2 + v_r t$, 其中, $a_r = a_i + a_j$, 并且 $v_r = v_i + v_j$. 将其带入式(9),得到最大的链路维持时间 t 为

$$t = \frac{-v_r + \sqrt{v_r^2 + 2a_r(R - X)}}{a_r} \quad (10)$$

而当两个车辆同向而行时,由于存在加速、减速、超车的现象,对于判断链路状态,我们判断车辆 i 与 j 哪一

个在前是至关重要的. 当 $S_j(t) - S_i(t) + X > 0$ 时,车辆 j 位于车辆 i 的前面,反之车辆 i 超过车辆 j ,位于车辆 j 的前面. 为了能够有效地表示出哪一个车辆在前,我们定义一个符号函数,如下

$$I(i, j) = \begin{cases} 1, & \text{满足 } S_j(t) - S_i(t) + X > 0 \\ -1, & \text{其他} \end{cases} \quad (11)$$

当链路处于断开的临界状态时满足如下公式

$$S_j(t) - S_i(t) + X = R \cdot I(i, j) \quad (12)$$

此时,我们分为以下 2 种情况来计算链路的维持时间.

当 $I(i, j) = 1$ 时,此时车辆 j 位于车辆 i 的前面,由式(12)可知, $S_j(t) - S_i(t) + X = R$, 与式(10)的计算类似,此时由于 $S_j(t) - S_i(t) = \frac{1}{2} \cdot a_r t^2 + v_r t$, 其中 $a_r = a_j - a_i, v_r = v_j - v_i$,我们能够得出时间 t 为

$$t = \frac{-v_r + \sqrt{v_r^2 + 2a_r(R - X)}}{a_r} \quad (13)$$

同理,当 $I(i, j) = -1$ 时,此时车辆 i 位于车辆 j 的前面,由式(13)得到 $S_j(t) - S_i(t) + X = -R$, 计算链路维持时间 t 为

$$t = \frac{-v_r - \sqrt{v_r^2 - 2a_r(R + X)}}{a_r} \quad (14)$$

由式(10)(13)(14)能够计算出发送节点与任意节点在一跳范围内的链路维持时间.

3 RSAR 路由算法

3.1 VANET 中的 Q-Learning 策略

标准的 Q-Learning 是一种基于学习主体 Agent 进行的启发式学习方式. 一般来说,在标准 Q-Learning 算法中, Agent 的学习过程主要由一个 3 元组组成 $\{S, A, R\}$. 其中, $S = \{s_1, s_2, \dots, s_n\}$ 表示 Agent 的状态空间, $A = \{a_1, a_2, \dots, a_n\}$ 表示 Agent 的活动空间, 从一个状态转移到另一个状态被视为一次有效的活动, R 代表 Agent 进行一次活动时所获得的即时奖励值. 越接近于目的地, 则其进行一次活动获得的奖励越高. 这里先给出几个相关定义.

定义 1 基本组件

学习环境 G : 将整个 VANETs 环境作为 Agent 的学习环境.

主体 Agent: 学习主体为每一个车辆节点.

状态空间 S : 本 Agent 外的其他节点组成本 Agent 的状态空间.

活动空间 A : 将信标数据包从一个车辆转发到另一个车辆则定义为一次活动.

即时奖励 R : Agent 一次活动获得的即时奖励.

定义 2 奖励值 主体 Agent 进行一次活动所获得

所得到的值称之为奖励值(即:Q-Value 值 $Q(s, a)$ ($s \in S, a \in A$) 来表示),其范围从 $[0, 1]$. 因为距离目的节点一跳的邻居节点能够直接到达目的节点,所以将其奖励值设为 1. 由式(15)定义整个网络中的初始奖励值为固定值 R

$$R = \begin{cases} 1, & \text{if } s \in N_d \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

其中, N_d 表示目的节点 d 的一跳邻居节点集合. 目的节点所有邻居节点一次活动获得的奖励值为 1.

定义 3 Q-Table 表 每一个学习主体 Agent 维护着一张二维表,用来记录其所能够到达的目的节点地址以及与其一跳邻居节点的 Q-Value 值,称这张二维表为 Q-Table 表(如表 1 所示).

表 1 Q-Table 表

Neigs	Dests		
	D_1	D_2	...
N_1	$Q(D_1, N_1)$	$Q(D_2, N_1)$...
N_2	$Q(D_1, N_2)$	$Q(D_2, N_2)$...
...

表中第一行 Dests 代表本节点能到达的目的节点表示 ID,在这里用 D_i 表示. 第一列 Neigs 表示与其相邻的一跳邻居节点表示 ID,在这里用 N_i 表示. $Q(D_1, N_1)$ 表示本节点到达目的节点 D_1 时与邻居节点 N_1 之间的 Q-Value 值. 如表 1 所示, Q-Table 表是一个二维表的形式,其大小由邻居节点数目以及在学习环境 G 中目的节点个数来决定. 容易看出,其有着很好的可扩展性. 节点间通过周期性的交换信标数据包来更新 Q-Table 的值. 将学习的任务分布到每一个节点中,使得算法能够快速收敛到最优路径,而且对网络拓扑的变化能够做出及时的调整.

在上面的定义中,每一个车辆节点被定义为一个学习主体 Agent. 与传统的学习方式不同的是,每一个车辆节点中存储有一张 Q-Table 表,节点间通过交换信标信息来更新各自的 Q-Table 表来完成学习的任务. 当一个节点发出一个信标数据包时,在信标数据包中不仅包含自身的速度,位置等信息,而且还包含了到达某一目的地址其与邻居节点的最大 Q-Value 值. 也就是说某一列中的最大值(如表 1 所示). 假设有如图 2 所示的 VANETs 网络拓扑环境 $G = \{V, E\}$. 其中, $V = \{A, B, C, \dots, H\}$ 表示车辆节点集,对节点 A 来说,其状态空间 S_A 为不包含 A 的所有节点组成的集合. 边集 E 表示所有能够在一跳范围内直接通信的节点连线组成的集合. 假设在图 2 中, A 为源节点, G 为目的节点. 现在要通过自学习的方式找出一条从发送节点 A 到目的节点 G 的最佳路径.

因为学习任务被分配到了每一个车辆节点中,即

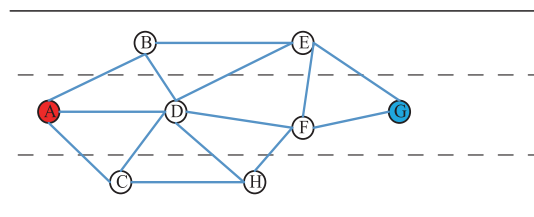


图 2 学习过程模型

每一个车辆被视作 Agent,而学习的过程主要是去更新 Agent 中的 Q-Table 表,也就是去更新 Q-Value 的状态活动对 $Q(s, a)$ ($s \in S, a \in A$). 标准的 Q-Learning 学习函数如式(16)所示

$$Q_s(d, x) \leftarrow Q_s(d, x) + \{R + \gamma \cdot \max_{y \in N_x} Q_x(d, y)\} \quad (16)$$

其中, $Q_s(d, x)$ 为要更新的 Q-Value 值. 其中 s 为本点, x 为其邻居节点, N_x 为 x 的邻居节点, d 为目的节点, R 为奖励值, $\max_{y \in N_x} Q_x(d, y)$ 为到达目的节点 d , x 与其邻居节点的最大 Q-Value 值. 其中,打折因子 γ 是一个重要的参数,根据以上公式我们知道其影响本节点进行一次活动获得的奖励值. 在 VANETs 网络中,可利用的带宽值作为一个重要参数,其决定着数据包传输的速率. 带宽 BW 计算公式可定义如下

$$BW(\text{bps}) = \frac{n \times S_b \times 8}{T} \quad (17)$$

其中, n 表示节点发出和接收到的数据包的数目. S_b 是数据包的大小,用字节表示. T 为时间间隔. 假设节点的最大可利用带宽值为一固定值,这里定义为 \max_{BW} , 发送信标数据包所占用的带宽大小为 BW_{beacon} , 计算公式为式(17),那么可以计算出有效的带宽因子为

$$BF = \frac{\max_{BW} - BW_{\text{beacon}}}{\max_{BW}} \quad (18)$$

将带宽因子作为影响学习快慢的因素,随着有效带宽的变化,其决定每一个车辆节点的学习进度. 我们修改式(16)得到新的启发式函数如式(19)

$$Q_s(d, x) \leftarrow (1 - BF) \cdot Q_s(d, x) + BF \cdot \{R + \gamma \cdot \max_{y \in N_x} Q_x(d, y)\} \quad (19)$$

每一个节点利用式(19)进行迭代计算,距离目的节点跳数越多的节点其最终得到的 Q-Value 奖励值将越低. 所以最终的奖励值是根据跳数、链路可靠性以及带宽这 3 个因子来决定. 通过加入带宽以及链路状态这两个参数,使得最终能够在这种动态的网络中得到从源到目的节点的最优路径. 以一次寻路学习过程为例:每一个节点中的 Q-Table 表中值被初始为 0,假设在图 3 中, A 点发出的广播数据包已经到达了目的节点 G , 节点 E 和 F 为目的节点 G 的一跳邻居节点,根据定义 2,

得到节点 E 和 F 到目的地 G 的奖励值分别为 $Q_E(G, G)$ 和 $Q_F(G, G)$, 且大小为 1. 考虑到受链路质量以及带宽的影响, 我们假设其最终得到的 Q-Value 值分别为 0.7 和 0.8. 在广播过程中, A 到 D 之间的所有节点会更新目的地址行 (Q-table 的第一行), 同时通过不断接收邻居节点发来的信标来更新邻居节点列 (Q-table 的第一列). 在 Q-Table 表中, 目的节点不为空的 Agent 节点会不断的利用式 (19) 去更新 Q-Table 表, 直到收到源节点发来的取消通知来停止学习过程并删除数据项. 考虑 D 的邻居节点有 A, B, C, E, F, H , 当 D 收到其任何一个邻居节点发来的信标数据包时, 解析数据包, 从中抽取到达目的节点 G 的最大 Q-Value 值. 以中间节点 F 为例, 因为 $Q_F(G, F)$ 是可以确定的, 根据式 (19) 计算出 $D \rightarrow F$ 对应的 Q-Value 值, 即 $Q_D(G, F)$, 并更新节点 D 的 Q-Table 表中对应的值. 同理, 当节点 D 收到 E 发来的数据包时也做同样的处理, 并且更新 Q-Table 表. 依此类推, 收到其他节点发来的数据包时也做如上处理, 这样便更新了节点 D 的 Q-Table 表中目的地址为 G 的一系列数据. 考虑到受带宽以及链路可靠性的影响, 我们假设根据式 (19) 得到 $Q_D(G, F) = 0.5$, 与其他邻居节点更新的 Q-Table 值如图 3 所示. 随着不断收到邻居数据包, 节点 D 通过不断修改其 Q-Table 表中保存的其与邻居节点的最大 Q-Value 值. 当节点 D 要发送信标数据包时, 其遍历保存在 Q-Table 表中目的地址为 G 的列, 找出最大的 Q-Value 值记录到信标数据包中发送出去, 当节点 A 收到其邻居节点 D 发来的信标数据包时, 利用式 (19) 计算出到目的节点 G 的 $A \rightarrow D$ 的最大 Q-Value 值, 并进行更新 A 的 Q-Table 表项, 收到其他节点发来的信标数据包也做同样的处理. 通过不断的进行数据包的交换, 最终我们将得到如图 3 所示的结果.

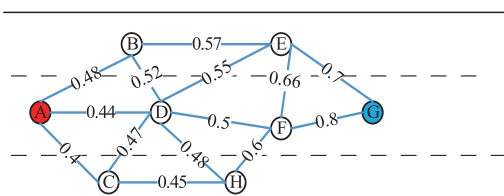


图3 每一个节点中保存的到达目的节点 G 的 Q-Table 值

根据图 3 的运行结果, 当 A 要发送数据包到 G 时, 只需要找出具有最大 Q-Value 值的邻居节点组成的路径即为最优路径. 从图 3 中可以看出, $A \rightarrow B \rightarrow E \rightarrow G$ 为具有最大 Q-Value 值的路径. 每一个节点动态更新自己的 Q-Table 表, 使得算法对拓扑的动态变化能够做出即时的响应, 在保证可靠性的同时, 也具有很好的鲁棒性.

3.2 RSAR 路由算法描述

RSAR 路由算法转发数据包的过程如下.

(1) 当源节点要发送一个数据包时首先去查找自己的 Q-Table 表, 查看是否有到达目的地址的下一跳节点, 如果有, 则从中选出具有最大 Q-Value 值的邻居节点, 否则启动路由建立过程, 路由建立过程见 3.2.1 节.

(2) 经过路由建立过程后, 得到了一条从源到目的节点的基本路径, 并且完成了对部分车辆节点的学习. 为了能够找出整个网络拓扑中的最优路径以及解决网络分割问题, 启动路由维护过程, 动态的维护端到端的路径, 路由维护过程见 3.2.2 节.

(3) 经过上面 2 个过程能够建立整个拓扑网络中的最优路径. 当某一个车辆点收到/发送数据包时, 执行第 1 步, 否则执行第 2 步.

3.2.1 路由建立过程

在开始阶段, 当源节点要发送一个数据包给目的节点时, 去查找 Q-Table 表, 查看是否有到达目的节点的下一跳节点. 如果有, 则从 Q-Table 表中找出一个到达目的节点最大的 Q-Value 值的邻居节点, 并将数据包转发给它. 如果不存在, 则启动路由发现过程. 源节点以广播的形式发向整个网络发送一个 R_req 数据包, 并启动路径请求定时器. 其中 R_req 在路由过程中记录着所经过的节点 ID. 当目的节点收到第一个来自源节点的 R_req 数据包时, 将其保存, 并丢弃随后收到的重复数据包. 通过解析接收到的 R_req 数据包, 目的节点将其中所记录的所有中继的 ID 节点信息抽取出来, 同时生成一个 R_req 数据包, 并将源节点作为目的节点, 使 R_req 按源路径逆向返回, 目的是去通知源节点, 同时达到首次更新中继节点的 Q-Table 表. 当记录中的节点收到此数据包时, 修改数据包的下一跳地址, 更新其 Q-Table 表, 并仍然以单跳广播的形式发送出去. 其他非目的节点收到时只修改 Q-Table 表, 然后丢弃. 依此类推, 直到 R_req 被发送到目的节点, 也就是 R_req 的源节点. 源节点取消请求定时器, 同时更新其 Q-Table 表. 这时, 便建立了一条从源节点到目的节点的路径, 同时也第一次更新了源节点到目的节点第一条路径上的节点中的 Q-Table 表, 初始的 Q-Table 表中的值为 0.

3.2.2 路由维护过程

当建立了第一条路由路径后, 与此同时也更新了与这条路径相邻的部分节点的 Q-Table 表中的值. 为了能够使得在动态变化的网络中保证路径的有效性, 启动路由维护过程是至关重要的. 路由维护过程主要目的是去动态的维护 Q-Table 表以及解决网络分割问题. 每一个节点周期性的去广播信标数据包来更新邻居节点的 Q-Table 表. 其中, 信标数据包主要由节点的位置信息、速度信息、 $\max(Q\text{-value})$. 为了能够保证更新的实效性, 在实验中设定信标数据包的发送时延为 $[0.5, 1]$ 之间的随机数. 其中, 为 Q-Table 表规定了目的节点的

有效时间. 如果某一个目的节点的时间由于没有得到更新而超过规定的时间, 则认为此目的节点失效, 并删除其对应的一系列数据. 当由于车辆的移动导致出现网络分割时, RSAR 在分割节点处采用 carry-and-forward 策略, 同时启动路径请求定时器去广播 R_{req} 数据包, 如果到预定时间没有收到目的节点发来的 R_{rep} 数据包时, 则认为目的节点不可达, 同时通知源节点取消数据包的传输, 否则从断裂处重新建立路由路径.

4 实验仿真

我们利用 NS-2 来进行实验仿真, 利用车辆运动生成器 (VanetMobiSim) 生成一个 1500×1500 的正方形拓扑场景, 并随机生成运动节点, 如图 4 所示 (因为实验分多步完成, 每次设置的节点个数不同, 这里仅仅给出道路模型). 道路由交叉路口以及直线路组成, 每一条道路均设置为 2 车道的双向车道, 在 3 个交叉路口处设置交通灯, 交通灯的变化时间为 5s. 为了突出真实性, 在整个网络环境中, 每一个车辆节点采用智能驾驶模型 (IDM), 其中包括变道、超车、避让以及等待, 每次实验节点的个数以及行驶的平均速度见表 2, 然后利用 VanetMobiSim 生成每一个节点移动的数据并加入到 NS-2 中进行仿真. 在 NS-2 中, 基本的网络参数配置见表 2.

表 2 网络参数配置

参数	值
拓扑大小 (m)	1500×1500
MAC 标准	IEEE 802.11 MAC (2Mbps)
传输范围 (m)	250
传播模型	two-ray ground
仿真时间 (s)	300
CBR 数据包大小 (byte)	512
数据率 (packet/s)	10
节点运动速度 (km/h)	30 ~ 90
节点个数	60 ~ 120

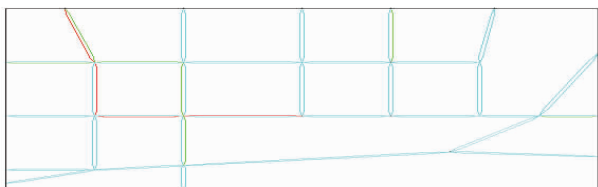


图 4 仿真拓扑图

在整个网络环境中, 我们随机生成 10 对 CBR 数据流来发送数据包, 数据包的大小为 512 字节, 传输层使用 UDP. 信标数据包的大小根据其所传输的内容字节进行计算. 为了能够有效的验证 RSAR 路由算法的性

能, 我们分别在图 4 所示的场景中, 设置多种情况来进行仿真. 例如: 设置整个网络中的节点数目固定为 80 个, 车辆的最大速度从 30km/h 到 90km/h, 在这一场景中, 主要研究车辆速度对路由协议的影响. 在实验开始时, 节点被随机的分散在不同的道路上, 并按固定的路线行驶. 每次的仿真时间为 300s, 每一种情况都进行 20 次仿真后取平均值, 得出仿真结果.

本文所提出的 RSAR 路由协议分别与 GPSR^[21] 算法, SLBF^[12] 算法以及 QLAODV^[13] 算法进行对比实验. 这 3 种算法在 VANETS 中具有代表性的路由算法, 通过与这 3 种路由算法对比, 来评估 RSAR 算法的优势以及不足. 在不同的场景模型中, 通过比较这 3 种算法的数据包的递交率, 传输延迟以及跳数. 我们得到了图 5 至图 6 的结果.

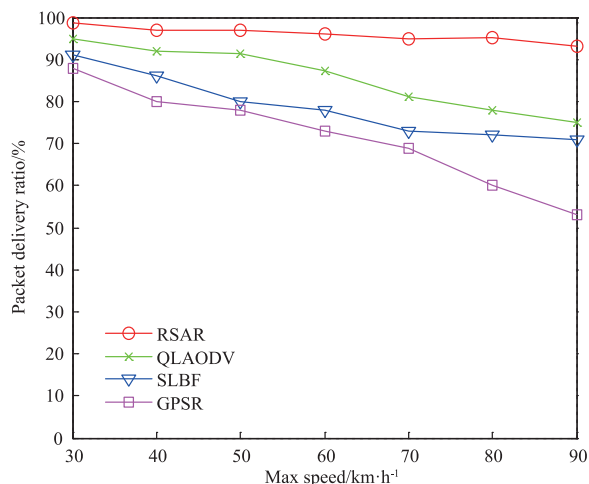


图 5 包的递交率与速度的关系

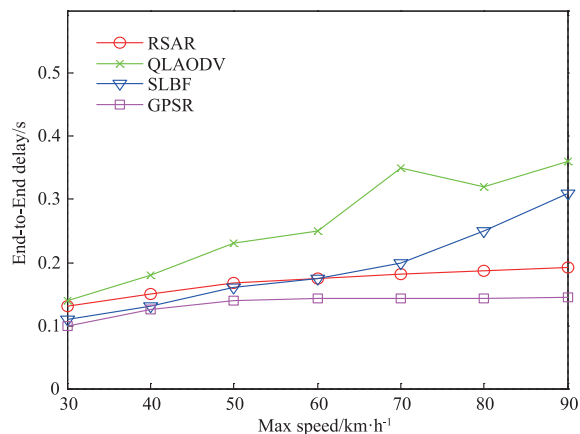


图 6 延迟时间与速度的关系

从图 5 中可以很明显的看出, 随着车辆速度的增加, 本文提出的 RSAR 表现出很高的包的递交率, 而且很稳定, 平均递交率达到 90% 以上. 而另外的 3 种路由算法随着车辆速度的增加, 包的递交率呈现急剧下降

趋势. 这是因为 RSAR 充分考虑了速度变化对链路稳定性的影响. 通过对节点间的链路评估, 确定出节点之间链路的可靠性, 并将其作为 Q-Learning 算法的一个学习参数, 应用在路由决策中. 从图 5 可以看出, GPSR 递交率下降最快, 这是由于其选择下一跳节点时没有考虑链路可靠性, 同时贪婪机制也是递交率下降的主要因素. 当速度小于 54km/h 时, QLAODV 的递交率也在 90% 以上, 但当速度大于 54km/h 时, 其递交率急剧下降. 这其实很容易理解, QLAODV 虽然采用了 Q-Learning 模型, 但是由于其要维护一条端到端的可靠路径, 使得随着速度增加, 使得其不断进行路径修复, 导致降低了包的递交率. SLBF 受拓扑变化的影响, 同时没能充分考虑链路的稳定性, 使得其递交率下降. 同时由于其也采用贪婪的方式, 使其更加接近于 GPSR 路由算法.

从图 6 中可以看出, 随着车辆节点运动速度的增加, 4 种算法的延时时间都呈现上升的趋势. 其中本文提出的 RSAR 处于 GPSR 与 SLBF 这两种路由算法之间, 而且比较稳定. 当最大速度大于 60km/h 时, SLBF 的延时时间迅速增大, 并且超过了 RSAR, 这是由于其受拓扑变化的影响, 重新计算有效转发区或重传机制所导致的. 而 RSAR 受拓扑影响比较小, 最大 Q-Value 值的路径必然有着最大的带宽及最可靠的链路和最短的路由长度, 所以, 能保证其有着最小延时时间. 从图中明显的可以看出, GPSR 路由算法的延时最小, 这是由于 GPSR 只采用了贪婪转发模式.

我们利用链路维持时间模型对 Q-Learning 学习函数进行了修改, 针对链路维持时间参数对性能的影响问题, 除了上述表中所给定的参考外, 我们还进行了大量的仿真实验和实际道路场景的测试与验证. 实验和测试表明, 链路维持时间参数对 Q-Learning 学习性能的影响呈现形如正态分布样式的非线性关系: 当 Q-Learning 学习过程未收敛的情况下, 链路维持时间越长, Q-Learning 学习的效果越好, 本文提出的 RSAR 算法性能也越好; 当 Q-Learning 学习过程达到收敛状态以后, 链路维持时间的长短对 RSAR 算法性能的影响呈现越来越小的关系(甚至影响为 0). 多组实验数据表和实验对比图略.

5 结束语

针对移动物(车)联网的路由问题, 通过分析车辆的运动特点, 我们研究了造成链路不可靠的因素, 证明了链路的维持时间服从于对数正态分布. 在其基础上建立了节点间链路可靠性计算模型. 将评估节点间链路的可靠值作为一个参数应用在 Q-Learning 学习策略中, 我们提出了新的路由算法. 通过实验对 4 种路由的性能(递交率、延时以及跳数)进行了评估比较. 实验结

果表明, 新算法在各种条件下均有着较高的包递交率和很低的传输延时, 能够有效地解决拓扑变化所带来的问题.

参考文献

- [1] Namboodiri V, Gao L. Prediction-based routing for vehicular ad hoc networks [J]. *IEEE Transactions on Vehicular Technology*, 2007, 56(4): 2332-2345.
- [2] Zhang D G. A new approach and system for attentive mobile learning based on seamless migration [J]. *Applied Intelligence*, 2012, 36(1): 75-89.
- [3] Liu S. Novel unequal clustering routing protocol considering energy balancing based on network partition & distance for mobile education [J]. *Journal of Network and Computer Applications*, 2017, 88(15): 1-9.
- [4] Ma Z. A novel compressive sensing method based on SVD sparse random measurement matrix in wireless sensor network [J]. *Engineering Computations*, 2016, 33(8): 2448-2462.
- [5] Zhang X D. Design and implementation of embedded un-interruptible power supply system (EUPSS) for web-based mobile application [J]. *Enterprise Information Systems*, 2012, 6(4): 473-489.
- [6] Song X D. Extended AODV routing method based on distributed minimum transmission (DMT) for WSN [J]. *International Journal of Electronics and Communications*, 2015, 69(1): 371-381.
- [7] Zhao C P. A new medium access control protocol based on perceived data reliability and spatial correlation in wireless sensor network [J]. *Computers & Electrical Engineering*, 2012, 38(3): 694-702.
- [8] Zhang D G, Liang Y P. A kind of novel method of service-aware computing for uncertain mobile applications [J]. *Mathematical and Computer Modelling*, 2013, 57(3-4): 344-356.
- [9] Zhang D G, Li G. An energy-balanced routing method based on forward-aware factor for wireless sensor network [J]. *IEEE Transactions on Industrial Informatics*, 2014, 10(1): 766-773.
- [10] Ma Z. Shadow detection of moving objects based on multisource information in internet of things [J]. *Journal of Experimental & Theoretical Artificial Intelligence*, 2017, 29(3): 649-661.
- [11] Zhang D G. A novel image de-noising method based on spherical coordinates system [J]. *EURASIP Journal on Advances in Signal Processing*, 2012, 2012(110): 1-10. DOI: 10.1186/1687-6180-2012-110.
- [12] Eiza M H. An evolving graph-based reliable routing scheme for VANETs [J]. *IEEE Transactions on Vehicu-*

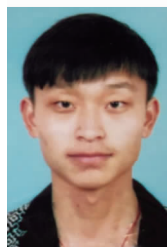
- lar Technology, 2013, 62(4):1493 - 1504.
- [13] Yan G J, Olariu S. A probabilistic analysis of link duration in vehicular ad hoc networks [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12 (5): 41227 - 41236.
- [14] 张德干. 基于局域世界的 WSN 拓扑加权演化模型 [J]. 电子学报, 2012, 40(5):1000 - 1004.
ZHANG De-gan. Local-world weighted topology evolving model for wireless sensor networks [J]. Acta Electronica Sinica, 2012, 40(5):1000 - 1004. (in Chinese)
- [15] 钟远. 一种不可信环境下的匿名位置辅助路由激励机制 [J]. 电子学报, 2013, 41(3):475 - 480.
ZHONG Yuan. An anonymous incentive mechanism for location-aided routing in untrusted MANET scenarios [J]. Acta Electronica Sinica, 2013, 41(3):475 - 480. (in Chinese)
- [16] 张德干. 一种基于前向感知因子的 WSN 能量均衡路由方法 [J]. 电子学报, 2014, 42(1):113 - 118.
ZHANG De-gan. A kind of routing protocol based on forward-aware factor for energy balance [J]. Acta Electronica Sinica, 2014, 42(1):113 - 118. (in Chinese)
- [17] Zhang D G. A novel approach to mapped correlation of ID for RFID anti-collision [J]. IEEE Transactions on Services Computing, 2014, 7(4):741 - 748.
- [18] Zhu Y N. A new constructing approach for a weighted topology of wireless sensor networks based on local-world theory for the internet of things (IOT) [J]. Computers & Mathematics with Applications, 2012, 64 (5): 1044 - 1055.
- [19] Song X D. New agent-based proactive migration method and system for big data environment (BDE) [J]. Engineering Computations, 2015, 32(8):2443 - 2466.
- [20] Zhang D G. A kind of novel VPF-based energy-balanced routing strategy for wireless mesh network [J]. International Journal of Communication Systems, 2017, 30(6):1 - 15.
- [21] Wang X. New clustering routing method based on PECE for WSN [J]. EURASIP Journal on Wireless Communications and Networking, 2015, 2015(162):1 - 13. DOI:10.1186/s13638-015-0399-x.
- [22] Zheng K. Novel quick start (QS) method for optimization of TCP [J]. Wireless Networks, 2016, 22(1):211 - 222.

- [23] Zheng K. A novel multicast routing method with minimum transmission for WSN of cloud computing service [J]. Soft Computing, 2015, 19(7):1817 - 1827.

作者简介



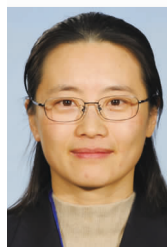
张德干 男, 1970 年 1 月出生, 湖北黄冈英山县人. 博士, 教授, 博导. 主要研究方向为物联网、移动计算、网络通信、智能控制等.
E-mail: zhangdegan@tsinghua.org.cn



葛辉 男, 1993 年 3 月出生, 山东济南人. 天津理工大学硕士研究生. 主要研究方向为物联网、网络通信等.
E-mail: 1464736574@qq.com



刘晓欢 (通信作者) 女, 1989 年 6 月出生, 黑龙江大庆人. 天津理工大学博士研究生. 主要研究方向为物联网、网络通信等.
E-mail: 815215568@qq.com



张晓丹 女, 1972 年 5 月出生, 吉林通化人. 中国科学技术信息研究所研究员. 研究方向为机器学习等.
E-mail: 183650594@qq.com



李文斌 男, 1991 年 6 月出生, 天津市人. 天津理工大学硕士研究生. 主要研究方向为网络通信.
E-mail: 549411835@qq.com