

基于二部图匹配的车载网络分布式存储机制*

唐晓岚¹, 洪东惠¹, 陈文龙¹, 蒲菊华²

¹(首都师范大学 信息工程学院, 北京 100048)

²(深圳北航新兴产业技术研究院, 广东 深圳 518057)

通讯作者: 陈文龙, E-mail: chenwenlong@cnu.edu.cn



摘要: 现有的车载网络中对数据存储机制的研究大多以移动车载节点作为数据载体,然而车载节点的快速移动、存储空间有限、存在安全风险等特性,限制了车载网络数据存储性能的进一步优化.针对部署有路边基础设施的车载网络场景,以路边单元作为存储节点,提出了基于二部图匹配的车载网络分布式存储机制(distributed storage scheme,简称 DSS).在车载网络中,以最大化数据响应率为目标,路边单元的数据存储问题是 NP 完全问题.首先,依据请求分割规则将原问题转化为二部图最大匹配问题,其中,二部图左顶点代表车载节点请求,右顶点代表路边单元的存储单元;进而,利用 Hungarian 算法在多项式时间内求得最优解.由于问题转化可能造成不同路边单元存储相同数据的冗余问题,设计了冗余副本清理算法,依据不同副本的响应因子排序,检查并清理冗余副本.实验结果表明:DSS 能够提高数据响应率,降低响应时延,并保持较小的网络资源开销.

关键词: 车载网络;分布式存储机制;二部图匹配;冗余副本清理

中图法分类号: TP393

中文引用格式: 唐晓岚,洪东惠,陈文龙,蒲菊华.基于二部图匹配的车载网络分布式存储机制.软件学报,2016,27(9): 2377-2388. <http://www.jos.org.cn/1000-9825/5065.htm>

英文引用格式: Tang XL, Hong DH, Chen WL, Pu JH. Distributed storage scheme using bipartite graph matching for vehicular networks. Ruan Jian Xue Bao/Journal of Software, 2016, 27(9): 2377-2388 (in Chinese). <http://www.jos.org.cn/1000-9825/5065.htm>

Distributed Storage Scheme Using Bipartite Graph Matching for Vehicular Networks

TANG Xiao-Lan¹, HONG Dong-Hui¹, CHEN Wen-Long¹, PU Ju-Hua²

¹(College of Information Engineering, Capital Normal University, Beijing 100048, China)

²(Research Institute of Beihang University in Shenzhen, Shenzhen 518057, China)

Abstract: The vast majority of existing studies on data distribution in vehicular networks utilize the mobile vehicular nodes as content carriers, however the high mobility, the limited resources and the security risks of vehicular nodes prohibit the performance improvements. This paper proposes a distributed storage scheme named DSS using bipartite graph matching for vehicular networks where roadside units are regarded as storage devices. Considering the content replication at roadside units is NP-complete, DSS tackles the problem by

* 基金项目: 国家自然科学基金(61502320, 61373161, 61173009); 国家科技支撑计划(2014BAF07B03); 北京市教委项目(KM201410028015); 深圳市基础研究计划(JCYJ20140509150917445); 软件开发环境国家重点实验室开放课题(SKLSDE-2015ZX-25); 中央高校基础科研业务费; 北京市优秀人才培养资助青年骨干个人项目(2014000020124G133); 首都师范大学青年燕京学者培育对象

Foundation item: National Natural Science Foundation of China (61502320, 61373161, 61173009); National Key Technology Support Program (2014BAF07B03); Science & Technology Project of Beijing Municipal Commission of Education (KM201410028015); Science Foundation of Shenzhen City (JCYJ20140509150917445); State Key Laboratory of Software Development Environment (SKLSDE-2015ZX-25); Fundamental Research Funds for the Central Universities; Youth Backbone Project of Beijing Outstanding Talent Training Project (2014000020124G133); Cultivation Object of Young Yanjing Scholar of Capital Normal University

收稿时间: 2015-11-16; 采用时间: 2016-03-30

transforming it into a maximal matching problem of bipartite graph in which the left vertices stand for the requests from vehicular nodes and the right vertices represent the storage cells of roadside units. Hence the original problem is solved by Hungarian algorithm in polynomial time. In addition, a redundant content deletion algorithm overcomes the possible duplications generated from the problem transformation by checking the redundancy in the order of the response factors of different replicas. Experiments show that the DSS scheme achieves a high access ratio and keeps a short access delay with a small cost of network resources.

Key words: vehicular network; distributed storage scheme; bipartite graph matching; redundant content deletion

车载网络是由移动的车辆和固定的路边基础设施组成的车路协同系统^[1],通过对交通相关数据的检测、处理、存储和传输,利用车辆与基础设施(vehicle-to-infrastructure,简称 V2I)和车辆与车辆(vehicle-to-vehicle,简称 V2V)之间的通信,提供泛在的网络接入,为智能交通系统的多种应用提供技术支持,例如安全驾驶、路况监测、高速公路自动收费、多媒体资源共享等等^[2,3].

典型的车载网络如图 1 所示,其中, V_1, V_2, V_3 表示安装在移动车辆上的车载节点,车载节点携带传感设备,开展环境检测,同时,利用无线通信模块实现 V2V 数据交换^[4]. U_1 是路边单元,拥有比车载节点更强的计算和存储能力,通过有线链路或远距离高带宽无线链路提供可靠的网络连接,因此,路边单元常被用作互联网接入点^[5].车载网络支持 V2V 和车辆与路边单元(vehicle-to-roadside-unit,简称 V2U)通信,美国联邦通信委员会为车载通信制定了专门的标准,即,专用短程通信标准(dedicated short-range communications standards,简称 DSRC).

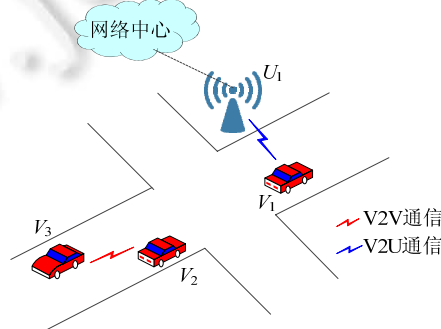


Fig.1 A sketch map of a typical vehicular network

图 1 车载网络示例

近年来,网络领域的部分研究工作转向以数据为中心,探索对数据的访问而非对主机的访问^[5],这一趋势亦影响到车载网络.车载网络通过车载节点为驾驶员和乘客提供服务,随着车载节点从网络获取数据的需求不断增长,网络资源匮乏和接入拥塞等问题愈加严峻.如何提升数据请求的响应率,是研究的重点与难点.已有研究工作大都利用移动的车载节点来存储数据,但考虑到车载节点移动速度快且行驶轨迹复杂,车载节点之间的通信存在较高的安全风险,因此,车载节点携带数据的成本高,稳定性差.

在车路协同系统中,固定部署的路边单元的通信能力、计算能力、存储能力优于车载节点,因此,本文提出一种以路边单元作为存储设备的分布式存储机制,旨在通过问题转化与建模来优化数据存储,提高数据响应率.具体而言,路边单元依据车辆的数据请求、自身的存储空间大小以及车辆与路边单元的预期相遇信息,决定其应当携带的数据;进而从接入互联网的控制中心获取并存储这些数据;当发起请求的车辆行驶到其通信范围时,通过 V2U 通信将数据传输给车载节点,完成数据响应.

本文的主要贡献在于:

- (1) 提出了请求分割规则,在此基础上,将数据存储的最优化问题(NP 完全问题)转化为二部图最大匹配问题,从而能够在多项式时间内求解;
- (2) 设计了冗余副本清理算法,旨在清除在问题转化过程中可能产生的冗余数据副本,保证网络资源的充分利用;

(3) 实验结果表明:与传统的数据存储方案相比,DSS 在使用较少资源的同时,有效地提高了数据响应率。

本文第 1 节介绍相关工作.第 2 节分析网络场景并进行问题建模.第 3 节详细介绍分布式存储机制 DSS 的核心组成部分,即,基于二部图最大匹配的分布式存储机制和冗余副本清理算法.第 4 节展示并分析实验结果.第 5 节总结全文.

1 相关工作

由于车载网络采用存储-携带-转发的数据传输模型,因此车载网络可看作机会网络的一种特殊应用.在机会网络中,传统的数据存储方案通常与路由策略一同讨论,典型的存储机制包括:

- (1) Epidemic^[7].当节点相遇时,每个节点均将其携带的数据发送给相遇的其他节点;每个节点都对所有数据进行存储;
- (2) Spray & Wait^[8].为防止信息拥塞,网络中每个数据的副本数是有限的;当节点相遇时,节点将其携带的一半副本发送给邻居节点,直到该节点只剩余一个副本.

上述方法实现简单,但数据存储代价大,传输效率受限.

近年来,一些研究侧重于设计节点协作的存储机制.在文献[9]中,若干个移动节点组成数据复制组,同一组中的各个车载节点分别存储服务器的一个数据子集,并以较低的通信成本与组内其他节点共享这些数据,从而提高网络的数据接收率.文献[10]设计了基于 P2P(peer to peer)通信原理的优化框架来管理数据复制和数据传播,该框架应用于没有固定基础设施的网络场景,选择车流作为虚拟主干网,解决了车辆之间数据传输的带宽限制问题.文献[11]研究了随机移动模型下的静态数据分布,探索最佳的数据分布模式,并具体分析了系统存储空间、数据大小和车辆行驶时间等网络参数的影响.对于车载网络的视频流应用,文献[12]提出了基于信任的概率存储机制,采用树型拓扑描述数据从源节点到目的节点的传输过程,并设计数据副本的成本函数和信任度计算方案.实验结果表明:当链路具有较高的稳定性且车辆移动速度较慢时,数据副本的成本降低;当存储空间大小和链路的可靠性增加时,数据传输速率上升.

目前,对于车载网络中分布式存储机制的研究主要利用车载节点来存储数据,根据车辆的请求和移动轨迹,计算数据副本的数目并选择合适的携带节点,若这些携带节点在请求的有效时间内与请求节点相遇,则通过 V2V 通信将数据发送给相应的请求节点.考虑到现有数据存储方案以一个完整的数据作为复制单元的不足,文献[13]构建数据包级的存储机制,通过预测车载节点的相遇持续时间,设计集中式的数据分配方案,旨在更好地利用有限的存储空间和宝贵的通信机会.该机制将数据副本作为开放的资源,自适应网络连接和车辆请求的动态变化,能够改善接收时延和拥塞情况.但是该机制基于移动节点间的多跳数据传输实现,因此只适用于低速移动的网络.文献[14]分析了网络容量对数据存储机制的影响,构建了当前副本数量、副本数量极值和节点数据队列长度之间的关系模型,提出了容量限制的分布式存储机制(distributed capacity-constrained replication scheme,简称 CCR),提高了数据接收率.

由于上述方案均由移动车载节点来复制和携带数据,其性能受到较大限制,因此,本文面向部署有路边单元的车载网络场景,使用存储资源更为丰富的路边单元来存储数据,以提高车载网络的数据服务质量.

2 问题建模

2.1 系统模型

本文讨论的车载网络场景中包括作为网络接入点的静态路边单元以及支持 V2V 和 V2U 通信(例如 IEEE 802.11p^[15])的动态车载节点,路边单元作为数据存储节点,而车载节点作为数据请求节点.具体地,以 $V=\{v_1, v_2, \dots, v_i\}$ 表示车载节点集合, $N_v=|V|$ 表示车载节点个数, $U=\{u_1, u_2, \dots, u_j\}$ 表示路边单元集合, $N_u=|U|$ 表示路边单元个数,任意路边单元 u_j 的存储空间大小记为 $r(j)$.车载节点请求的数据集为 $C=\{c_1, c_2, \dots, c_k\}$, $N_c=|C|$ 表示数据个数,任意数据 c_k 的大小为 $s(c_k)$.数据在路边单元中存储的有效时间为一个存储周期,存储周期的定义如下.

定义 1(存储周期 τ). 任意路边单元 $u_j(\forall u_j \in U)$ 从连接 Internet 的控制中心获取数据后,存储数据的有效时长为 τ ,称为存储周期.

在当前存储周期 τ 内,任意路边单元 u_j 将所存储的数据发送给相遇的请求车辆,并在数据失效时从控制中心下载在有效期内的新副本.

下面以车载节点 v_i 向路边单元 u_j 请求数据 c_k 为例,说明车载网络中的数据请求与响应过程.当 v_i 向 u_j 发起对 c_k 的请求时,若 u_j 携带数据 c_k ,则将 c_k 发送给 v_i ,车辆 v_i 接收数据则该请求满足;否则, v_i 继续行驶.在请求的有效期内与之相遇的其他路边单元若存储有数据 c_k ,则将 c_k 发送给 v_i ,响应数据请求;否则,请求失败.

假设车辆依据导航模块决定其在下一个存储周期内的行驶路线,进而获知将遇到的路边单元信息.当车辆与任意路边单元相遇时,通过 V2U 通信,路边单元获知车辆在下一个存储周期内将经过哪些路边单元,即,车载节点与路边单元的预期相遇信息.此信息将在本文提出的分布式存储机制中使用.

2.2 问题分析

考虑到车载节点和路边单元的存储能力有限,一个有效的数据存储机制应当综合利用二者的资源来提升服务质量,平衡网络流量.在密集的车载网络中,存储机制需要解决的关键问题包括:(1) 如何在路边单元中存储数据副本以尽可能多地响应请求;(2) 如何避免由于多个路边单元响应同一个请求造成的资源浪费.本文设计的分布式存储机制 DSS 将针对这两个问题分别给出解决方案.具体来讲:第 1 个问题针对路边单元的有限存储空间,当车辆发起数据请求时,考虑车辆与路边单元的相遇情况,决定将哪些数据副本存储在哪些路边单元,旨在尽可能多地响应车辆的请求;第 2 个问题针对存储在多个路边单元中的相同数据副本,通过冗余检测和消除,提升存储空间利用率,进一步优化车载网络对请求的响应.

为便于分析阐述,对网络参数做如下假设:任意数据请求的有效期为自请求发起时刻到下一个存储周期的结束时刻.所有数据大小相同,即:对于 $\forall c_k \in C$,满足 $s(c_k)=s$, s 的取值保证在车载节点和路边单元的一次通信过程中能够完成一个数据的传输.此外,任意路边单元 u_j 的存储空间大小是 s 的整数倍,即 $r(j)=w(j) \times s$,其中, $w(j) \in \mathbb{N}^+$,称为 u_j 有 $w(j)$ 个存储单元,每一个存储单元能够存储一个数据.

下文对分布式存储机制 DSS 的阐述基于以上假设,但是在实际应用中,通过简单调整能够将其应用范围扩大.例如:

- (1) 扩展到不同数据请求拥有不同有效期的场景.在 DSS 的每一个存储周期,控制中心都会对路边单元进行数据分配,请求有效期较长的数据有可能多次被分配到相关路边单元中,从而保持存储状态,实现被请求数据跨越多个存储周期的长时间存储;
- (2) 支持不同数据拥有不同大小的场景.依据网络常用的数据包分组与重组技术,大数据可以被分解成若干个小数据,在 DSS 中一个大小为 $b \times s(b \in \mathbb{N}^+)$ 的数据将被分解为 b 个大小为 s 的数据块,从而转变为相同大小的数据进行存储分配.

3 分布式存储机制 DSS

下面详述本文提出的分布式存储机制 DSS.为了方便描述,以 $f(i)(f(i) \subseteq C)$ 表示在一个存储周期 τ 内,任意车载节点 v_i 所请求的数据集合; $g(i)(g(i) \subseteq U)$ 表示 v_i 在下一个存储周期 $\tau+1$ 内将遇到的路边单元的集合.当 v_i 请求数据时, v_i 将 $f(i)$ 和 $g(i)$ 发送给与其通信的路边单元;然后,控制中心通过路边单元收集车辆的请求信息,并求解相应的数据分配方案.

本文以 $d(j)(d(j) \subseteq C)$ 表示分配给路边单元 u_j 的数据集合,网络的整体分配结果表示为 $D=\{d(1),d(2),\dots,d(j)\}$.数据存储的目标是,最大限度的满足车辆的请求.将分配方案 D 能够响应的请求个数记为 $P(f,g,D)$,其计算方法如下:

$$P(f,g,D) = \sum_{i=1}^{N_i} \left| f(i) \cap \left(\bigcup_{u \in g(i)} d(u) \right) \right| \quad (1)$$

在车载网络的数据存储问题中,每一个数据副本以其能够响应的请求个数为价值,以其数据大小为权重,同一数据的不同副本的价值之间相互影响(一个数据在多个路边单元上存储的多个副本可能响应同一个请求).由此可知,数据分布式存储问题是一个带约束的 0-1 背包问题.因此,原始问题是 NP 完全问题,不能在多项式时间内求解,本文研究的出发点在于寻找多项式时间内的近似最优解.

本文首先提出请求分割规则,在此基础上,将原始问题转化为经典数学问题,实现多项式时间内求解;然后,针对请求分割规则产生的潜在资源浪费问题,设计冗余副本清理算法削弱其影响,从而提高数据请求的响应率.其中,请求分割规则如下:

规则 1. 将每一个车载节点所请求的每一个数据看作独立的数据,即,不存在多个车载节点请求同一个数据的情况,表示为 $f(1), f(2), \dots,$ 和 $f(i)$ 不存在交集,称为请求分割规则.

DSS 执行过程如图 2 所示.

- 首先,在请求分割规则的约束下,利用二部图匹配算法求解分配优化问题,即,构建初始二部图 $G=(L, R, E)$,其中,左顶点集 L 表示车辆的请求,右顶点集 R 表示路边单元的存储单元,依据 $f(i)$ 和 $g(i)$ 添加边集 E ;利用 Hungarian 算法计算最大匹配子图 $GB=(L, R, EB)$,路边单元根据边集 EB 存储数据副本;
- 然后清理请求分割规则带来的冗余副本,并依据清理得到的剩余空间以及尚未满足的请求,再次分配数据,直到路边单元的存储空间耗尽或者所有请求都已满足或者剩下的请求再也无法完成,最终返回数据分配方案.

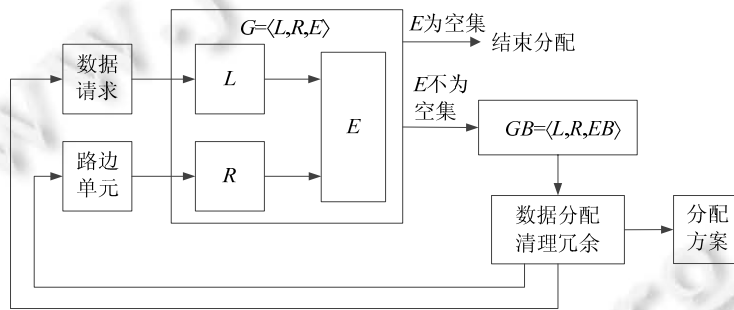


Fig.2 Process of the distributed storage scheme

图 2 分布式存储机制执行过程

举例详述上述过程,网络场景如图 3 所示.

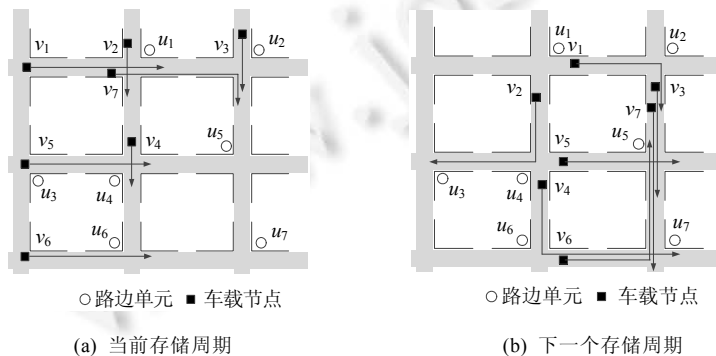


Fig.3 An instance scenario of distributed storage scheme

图 3 分布式存储实例的网络场景

车载网络中有 7 个车载节点(表示为 $V=\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$)和 7 个路边单元(表示为 $U=\{u_1, u_2, u_3, u_4, u_5, u_6,$

u_7 }).其中,路边单元的存储空间分别为 $w(1)=w(2)=w(3)=w(4)=w(6)=1, w(5)=3, w(7)=2$;在当前存储周期 τ 内,车载节点发出的数据请求分别为 $f(1)=\{c_1\}, f(2)=\{c_2\}, f(3)=\{c_1, c_3\}, f(4)=\{c_4\}, f(5)=\{c_5\}, f(6)=\{c_6\}, f(7)=\{c_1, c_7\}$.如图 3(b)所示:根据车载节点的预期行驶路线,在下一个存储周期 $\tau+1$ 内,车载节点将遇到的路边单元分别为 $g(1)=\{u_2\}, g(2)=\{u_4, u_3\}, g(3)=\{u_5\}, g(4)=\{u_6, u_7\}, g(5)=\{u_5\}, g(6)=\{u_7, u_5\}, g(7)=\{u_5, u_7\}$.

3.1 基于二部图最大匹配的分布式存储机制

依据车载节点的数据请求、路边单元的存储空间以及车辆与路边单元的预期相遇信息,控制中心构建一个初始二部图 $G=\langle L, R, E \rangle$.其中,任意左顶点 $L(i, k)$ 表示车载节点 v_i 请求数据 c_k ,任意右顶点 $R(j, z)$ 表示路边单元 u_j 的第 z 个存储单元.二部图的创建方式如下:对于任意数据 $c_k \in f(i)$,创建一个左顶点 $L(i, k)$;对于有 $w(j)$ 个存储单元的路边单元 u_j ,创建 $w(j)$ 个右顶点,分别表示为 $R(j, 1), R(j, 2), \dots, R(j, w(j))$;当且仅当 $u_j \in g(i)$ 时,在左顶点 $L(i, k)$ 和右顶点 $R(j, 1), R(j, 2), \dots, R(j, w(j))$ 之间分别创建一条边.图 3 中,实例的初始二部图如图 4(a)所示.

对于初始二部图 $G=\langle L, R, E \rangle$ 的一种匹配,若左顶点 $L(i, k)$ 和右顶点 $R(j, z)$ 之间存在一条边,则将数据 c_k 的一个副本存储在 u_j 的第 z 个存储单元.因此在分布式存储机制 DSS 中, $G=\langle L, R, E \rangle$ 的一个匹配能够代表一种存储方案.在请求分割规则的约束下,匹配中的一条边表示复制一个数据从而满足一个请求,即:匹配中边的数目反映了响应的请求个数,边越多,则满足的请求越多.因此,分布式存储问题转化成为寻找二部图最大匹配的问题.

二部图最大匹配问题是一个经典的图论问题, Hungarian 算法能够有效求解该问题.因此在 DSS 中,若初始二部图 $G=\langle L, R, E \rangle$ 的边集 E 不为空,则应用 Hungarian 算法求得最大匹配子图 $GB=\langle L, R, EB \rangle$;否则,分配过程结束.在图 3 实例中,由 Hungarian 算法求出的最大匹配子图如图 4(b)所示.

控制中心依据最大匹配子图 $GB=\langle L, R, EB \rangle$ 将数据副本分配到相应的路边单元,即:对于 EB 中任意一条连接左顶点 $L(i, k)$ 和右顶点 $R(j, z)$ 的边,控制中心生成数据 c_k 的副本并将其存储在路边单元 u_j 的第 z 个存储单元.由此得到初步的分布式存储方案,表示为 D_{imp}^B .

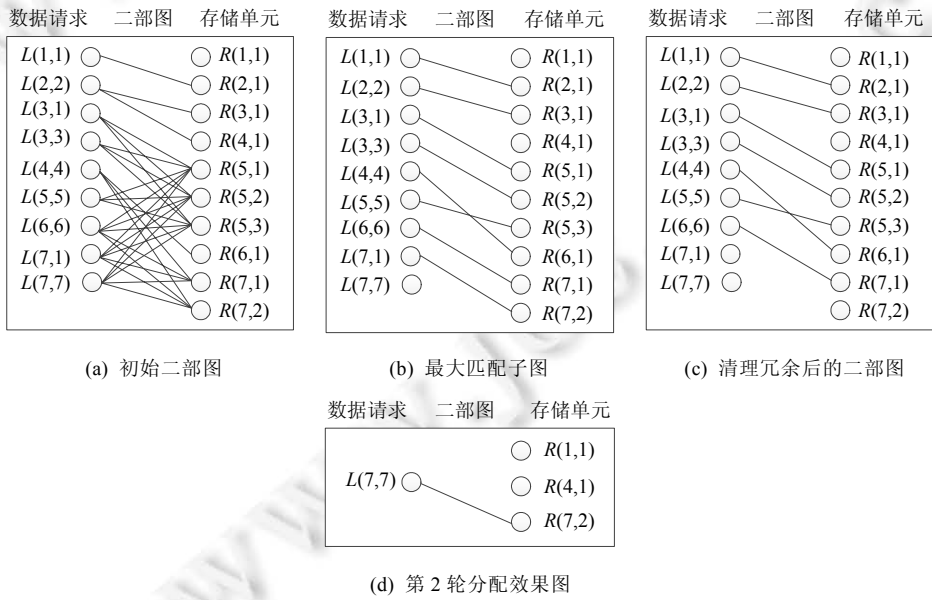


Fig.4 Construction and transformation of bipartite graphs in the distributed storage scheme instance

图 4 分布式存储实例的二部图创建与转变过程

3.2 冗余副本清理算法

考虑到在车载网络中多个车辆请求相同数据的情况很常见,因此需要去除请求分割规则的约束.由于初步

的分布式存储方案 D_{imp}^B 是基于请求分割规则得到的,若去除这一规则的约束, D_{imp}^B 中可能存在冗余数据副本.下面介绍两个典型的冗余存储案例.

- (1) 若车载节点 v_1 和 v_2 同时请求数据 c_1 ,为了满足 v_1 和 v_2 的请求,则 $u_1(u_1 \in g(1) \cap g(2))$ 会存储 c_1 的两份副本,造成冗余;
- (2) 若车载节点 v_1 和 v_2 同时请求数据 c_1 , u_1 存储 c_1 以响应 v_1 的请求,同时, u_2 也存储 c_1 以响应 v_2 的请求,此时,若满足 $(u_1 \cup u_2) \subseteq (g(1) \cap g(2))$,即, v_1 和 v_2 都将与 u_1 和 u_2 相遇,那么在 u_1 或 u_2 中存储一份副本即可.

因此,本文设计冗余副本清理算法,旨在避免路边单元的冗余存储,提高资源利用率.为解决前一种情况,可直接由各路边单元删除其携带的重复的数据副本.而在后一个案例中,情况较为复杂,需要控制中心逐一检查所有路边单元,查找并清理冗余副本,下面重点讨论此种情况.

若控制中心采用随机顺序对路边单元进行冗余检查,则可能无法清理所有冗余.例如在 D_{imp}^B 中, u_1, u_2 和 u_3 存储 c_1 ,在下一个存储周期 $\tau+1$ 中, u_1 将响应 v_1 和 v_2 的请求, u_2 将响应 v_3 ,而 u_3 将响应 v_1, v_2 和 v_3 .若冗余检查按照 u_1, u_2, u_3 的顺序进行,则检查结果为 u_1 和 u_2 中的副本不冗余, u_3 中的副本冗余而被删除.从资源利用的角度来看,这样的清理结果显然不是最优的,最优的结果应该是只保留 u_3 中的副本,而将 u_1 和 u_2 中的副本作为冗余删除.由此可见:为了尽可能多地清理冗余,应当参照每个副本响应的请求个数依次进行冗余检查.

冗余副本清理算法(算法 1)中, DU 表示存储 c_k 的路边单元集合; $NU(x)$ 表示能够从 $u_x(u_x \in DU)$ 接收 c_k 的车载节点集合; $|NU(x)|$ 表示 u_x 存储的 c_k 能够响应的请求个数,称为响应因子.

在冗余副本清理算法中,以任意数据 c_k 为例,首先将存储 c_k 的路边单元记入 DU ,若 DU 中只有一个元素,则表明存在车载节点请求该数据且在整个网络中只有一个副本,因此这个副本不冗余,不需要清理.在冗余检查过程中,控制中心将 u_x 能够响应的请求 c_k 的车辆记入 $NU(x)$.对于 DU 中的路边单元,按其携带的数据副本的响应因子降序排列,逐个进行冗余检查.

算法 1. 冗余副本清理算法.

Input: V, U, C, f, g, D_{imp}^B ;

Output: The distribution result after redundancy deletion D_{imp}^D .

```

1  for each data  $c_k \in C$  do
2    Initialize:  $DU \leftarrow \emptyset$ 
3    Initialize:  $NU \leftarrow \emptyset$ 
4    for each roadside unit  $u_j \in U$  do
5      if  $c_k \in d(j)$  then
6         $DU \leftarrow DU \cup \{u_j\}$ ;
9    if  $|DU| > 1$  then
10     for each roadside unit  $u_x \in DU$  do
11       for each vehicular node  $v_i \in V$  do
12         if  $c_k \in f(i)$  and  $u_x \in g(i)$  then
13            $NU(x) \leftarrow NU(x) \cup \{v_i\}$ ;
14       sort the roadside units in  $DU$  by descendent  $|NU(\cdot)|$ ;
15       for each roadside unit  $u_x \in DU$  in the above order do
16         if  $|NU(x)| > 0$  then
17           for each roadside unit  $u_y \in DU$  and  $u_y \neq u_x$  do
18             for each vehicular node  $v_i \in NU(y)$  do
19               if  $v_i \in NU(x)$  then
20                  $NU(y) \leftarrow NU(y) - \{v_i\}$ ;

```

```

21      $NU(x) \leftarrow \emptyset;$ 
22     else
23         delete the replica of  $c_k$  carried by  $u_x$ ;
24          $d(x) \leftarrow d(x) - \{c_k\};$ 
25 return  $D_{imp}^D$ ;

```

以任意路边单元 $u_x \in DU$ 为例,其冗余检查步骤如下:

检查 $NU(x)$ 中是否存在尚未满足的请求:若存在,则表明该副本不冗余,保留该副本,并将其他路边单元的 $NU(y)$ 中能够被 u_x 满足的请求删除;否则,即 $|NU(x)|=0$,表明 u_x 能够响应的请求已经被其他路边单元满足,则 u_x 中的 c_k 是冗余的,被清理。

清理冗余后得到更新的分布式存储方案,记为 D_{imp}^D 。在图3实例中,执行冗余副本清理算法, u_7 中的 c_1 被清理,得到清理后的二部图如图4(c)所示。

在冗余副本清理过程中,可能产生新的可用存储单元,因此,控制中心执行第2轮分布式存储,旨在充分利用存储资源来满足更多的请求。具体地,以尚未满足的请求为左顶点,以空余存储单元为右顶点,构建新的初始二部图,再次执行前述二部图最大匹配以及冗余副本清理的步骤,直到二部图无边,控制中心得到最终的分布式存储方案,记为 D 。在图3实例中,第2轮分布式存储将 c_7 分配到 u_7 的第2个存储单元,如图4(d)所示。

综上所述:在分布式存储机制 DSS 中,路边单元作为存储设备,周期性地从控制中心获取数据,并将副本发送给请求的车载节点。DSS 首先在请求分割规则的约束下,将原始问题转化成为二部图最大匹配问题,然后通过冗余副本清理算法解决规则造成的资源浪费,最终通过多轮数据分配,在多项式时间内计算得到一个响应率较高且资源利用充分的分布式存储方案。

4 实验结果及分析

4.1 网络环境配置

本文对 DSS 的实验验证在机会网络环境平台(opportunistic networking environment platform,简称 ONE^[16])上进行,ONE 被广泛认为是机会通信的标准模拟平台,虽然暂不支持诸如车流、变道等精确的交通模型,但 ONE 集成了多种典型的车辆移动模型和机会路由协议,这些组件简化了 DSS 中数据请求和响应过程的实现,因此,本文选择 ONE 平台进行模拟实验。

实验参数见表1。根据 IEEE 802.11p,车载节点和路边单元的通信半径是 200m,数据传输速率为 5Mbps。在基于地图的最短路径移动模型中,车辆从规定范围内随机选择速度,并以该速度沿着最短路径驶向目的地^[17]。路边单元的存储空间大小和车载节点的请求数量对服务质量的影响将在第4.3节中讨论。

Table 1 Simulation environment configurations

表1 仿真环境配置

参数	值
网络范围	网格状15000×15000m ² ,每个网格大小500×500m ²
路边单元数量, N_u	30×30,部署在路段交叉口
车载节点数量, N_v	1 000
仿真时长	6 000s
存储周期, τ	600s
通信半径	200m
数据传输速率	5Mbps
移动模型	基于地图的最短路径模型
速度范围	10~20m/s
数据数量, N_c	3 000
数据大小, s	10MB
路边单元存储空间大小	50MB
请求数目	在 t 时间内一个车载节点最多请求 5 个数据
车载节点请求数据	请求发起时刻随机,请求内容随机

由于经典的 Spray and Wait(S&W)方法和新近提出的 CCR 方法具有较好的分布式存储性能,本文选择这两种对比方法来测试 DSS 的效果.具体来讲,S&W 适用于机会网络,通过调节数据副本在移动节点中的存储来提升传输效率;而 CCR 是车载网络中容量限制的分布式存储机制,通过预测网络容量优化数据存储.由于车载网络是机会网络的典型应用,S&W 和 CCR 均适用于以车载节点为存储设备的车载环境.此外,考虑到针对部署有路边单元的车载网络存储机制的研究较少,选择时延优先的贪心算法作为二部图匹配算法的对比,称为 Greedy.在 Greedy 算法中,控制中心将车载节点 v_i 所请求的数据分配给它在下一个存储周期内将遇到的第 1 个路边单元($g(i)$ 中的第 1 个元素),若该路边单元的存储空间已满,则分配给 $g(i)$ 中的下一个路边单元.由于 Greedy 算法先将数据分配给较早相遇的路边单元,因此有助于缩短响应时延.

实验评估参照 4 个标准,即:数据响应率、平均响应时延、副本数量和传输开销.数据响应率表示被响应的数据请求在所有请求中所占的比率,响应率越高,则存储机制的性能越好.平均响应时延即获得响应的请求的响应时延的平均值,一个请求的响应时延是指从请求被发出到被响应之间的时间跨度,平均响应时延短则表示该机制适用于对时延有严格要求的车载应用.副本数量直接体现了整个网络中存储资源的消耗,而传输开销则表示在实验时间内传输的数据包个数,反映了网络通信资源的消耗.

4.2 实验结果及分析

在表 1 所示的实验参数下,各个分布式存储方案的性能如图 5 所示.

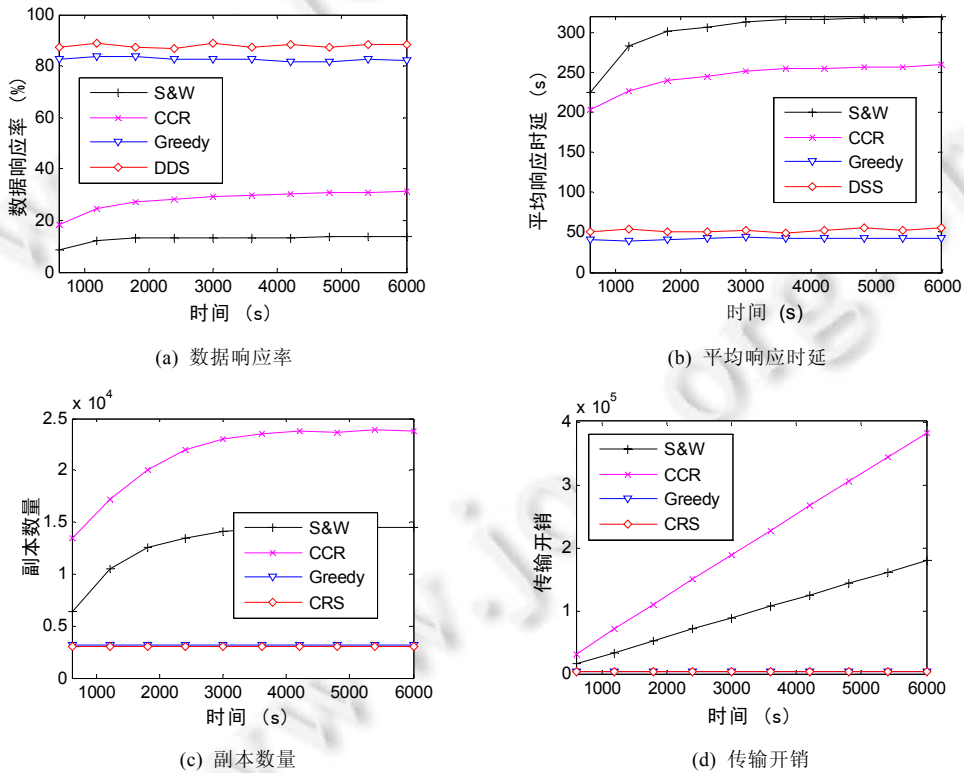


Fig.5 Experimental results

图 5 实验结果

在图 5(a)中,CCR 因为支持副本个数的动态调整,其响应率约是 S&W 的两倍.尽管如此,S&W 和 CCR 的数据响应率明显低于其他两种方案,这是由于 S&W 和 CCR 以车载节点作为存储设备,而车载节点之间的相遇时间短且相遇概率小,极大地限制了数据的传输.DSS 的响应率约为 88%,高于 Greedy.DSS 使用二部图最大匹配算

法来提升满足的请求数量,实验结果表明,DSS 的数据存储比 Greedy 更高效。

如图 5(b)所示:由于 Greedy 优先将数据分配给较早遇到的路边单元,故具有最短的平均响应时延,DSS 和 Greedy 的时延差距只有 10s,与 600s 的存储周期相比,DSS 和 Greedy 的时延差距几乎可以忽略不计。而 S&W 和 CCR 由于使用车载节点存储数据,时延远远大于其他两种方案。

图 5(c)和图 5(d)表明:S&W 和 CCR 的副本数量和传输开销远远高于 DSS 和 Greedy,其中,CCR 的副本数量最大。这是由于 S&W 中副本数量是不变的,而 CCR 会根据预测的网络容量来决定副本数量。此外,与 Greedy 相比,DSS 减少约 3%的副本数量,而两者的传输开销基本相同。

在本文提出的 DSS 机制中:

- 一方面,基于二部图最大匹配的分布式存储方法的计算复杂度与二部图节点数和边数相关,其时间复杂度为 $O(L \cdot |E|)$ 。由于 L 表示车辆的数据请求集合,假设在一个存储周期内任意车辆请求不多于 N_{rq} 个不同数据,则左顶点个数不大于 $N_{rq} \cdot N_v$;而边集 E 由车辆与路边单元的相遇情况决定,假设在一个存储周期内任意车辆最多遇到 N_{uc} 个不同路边单元,记路边单元的存储单元个数的最大值为 N_{us} ,即 $N_{us} = \text{MAX}(w(j))$,则边数不大于 $N_{rq} \cdot N_v \cdot N_{uc} \cdot N_{us}$ 。综合可得,该方法的时间复杂度表示为

$$O(N_{rq}^2 \cdot N_v^2 \cdot N_{uc} \cdot N_{us});$$

- 另一方面,由算法 1 可知,冗余副本清理算法的时间复杂度为 $O(N_c \cdot N_u^2 \cdot N_v)$ 。

综上所述,DSS 机制实现了在多项式时间内求解分布式存储问题。

整体来讲,与 CCR 机制相比,DSS 在大幅提高数据响应率的同时,将平均响应时延缩短至大约 20%;与贪心算法 Greedy 相比,DSS 的数据响应率提高了约 13%,而平均响应时延只是略有延长。总而言之,DSS 有效地提高了数据响应率,同时保持了较短的响应时延。

4.3 参数分析

本节讨论路边单元的存储空间大小和车辆的请求个数对存储机制性能的影响,为节省空间,此处仅分析数据响应率和平均响应时延这两项核心指标。

4.3.1 存储空间大小的影响分析

令路边单元的存储空间从 10MB~100MB 递增,考虑到 S&W 和 CCR 不使用路边单元,此处只比较 Greedy 和 DSS 的实验结果。如图 6 所示:随着存储空间扩大,Greedy 和 DSS 的响应率都先有所提高,然后平缓下降,同时响应时延不断缩短。主要原因是:(1) 存储资源匮乏限制了分布式存储方案 DSS 的优势发挥;(2) 当路边单元有足够的存储空间,就能存储更多的数据副本来满足请求并缩短响应时延;(3) 若路边单元的存储空间过大,则导致 V2U 通信时的带宽竞争加剧,最终使得响应率降低。

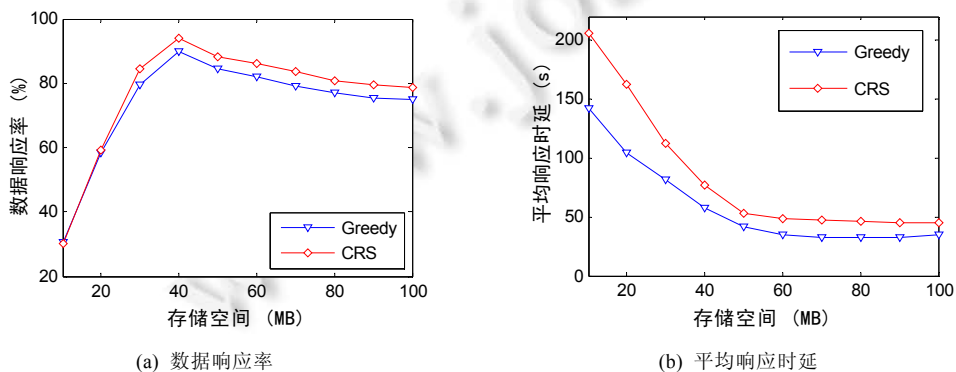


Fig.6 Results with different storage sizes of a roadside unit

图 6 路边单元存储空间大小的影响

4.3.2 请求个数的影响分析

令每个车载节点在一个存储周期内发送的请求个数从 1~10 依次递增,实验结果如图 7 所示.当请求不太多时,各个存储机制的性能基本保持稳定;在请求增加到一定程度后,网络的服务质量逐渐下降.车辆发送的请求越多,系统就需要越多资源来满足请求.因此,当请求个数适当时,整体性能保持不变.然而,当请求个数达到网络的饱和值,资源匮乏将极大地限制车载网络的整体性能.

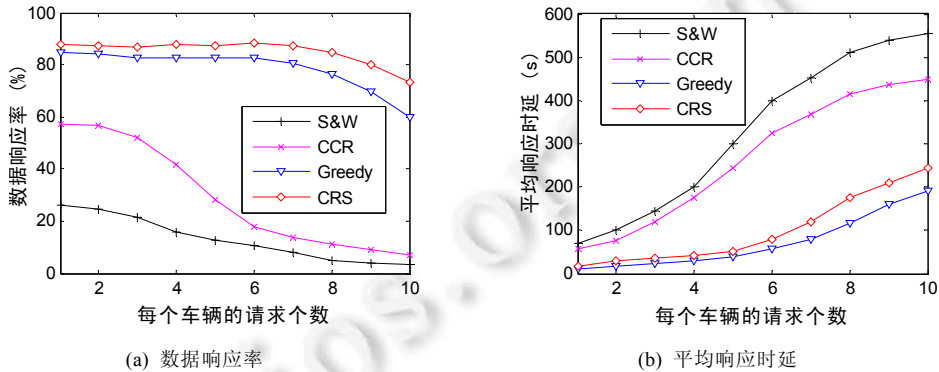


Fig.7 Results with different numbers of requests

图 7 请求个数的影响

综上所述,DSS 在不同的存储空间大小和不同的请求个数的条件下,都具有较高的响应率和较低的响应时延,进一步验证了其良好的可扩展性和通用性.

5 总结

本文提出了基于二部图匹配的分布式存储机制,以路边单元作为存储设备,旨在充分利用网络资源,有效地提高智慧交通应用中的数据接收率.首先设定请求分割规则,根据车载节点的数据请求、路边单元的存储空间大小以及车辆与路边单元的预期相遇情况,将初始的分布式存储问题转化为二部图最大匹配问题,在多项式时间内计算得到初始分配方案;然后设计冗余副本清理算法,针对不同路边单元存储同一数据时可能产生的冗余副本问题,依据响应因子对各副本排序,逐个进行冗余检查和清理,从而提高存储资源利用率.实验结果表明:与传统的车载网络存储机制相比,DSS 在有限的存储资源下,显著提高了数据响应率并保持较短的响应时延.

尽管如此,路边单元的无线信道竞争仍然是 DSS 实施的瓶颈,解决无线通信资源的调度问题将进一步提高整体性能^[18].此外,车载节点之间的协同下载^[19]有助于改善数据传输,将车辆协同下载技术与路边单元的分布式存储机制相结合,是未来的研究方向.

致谢 在此,我们向对本文的工作给予支持和建议的同行,尤其是北京航空航天大学计算机学院熊璋教授、蒲菊华副教授领导的讨论班上的同学和老师表示感谢.

References:

- [1] Piao J, McDonald M, Hounsell N. Cooperative vehicle in frastructure systems for improving driver information services: An analysis of COOPERS test results. *IET Intelligent Transport Systems*, 2012,6(1):9-17. [doi: 10.1049/iet-its.2010.0169]
- [2] Mershad K, Artail H. A framework for secure and efficient data acquisition in vehicular ad hoc networks. *IEEE Trans. on Vehicular Technology*, 2013,62(2):536-551. [doi: 10.1109/TVT.2012.2226613]
- [3] Wang SG, Lei T, Zhang LY, Hsu CH, Yang FC. Offloading mobile data traffic for QoS-aware service provision in vehicular cyber-physical systems. *Future Generation Computer Systems*, 2016,61:118-127. [doi: 10.1145/2491288.2491312]

- [4] Kenney JB. Dedicated short-range communications (DSRC) standards in the United States. Proc. of the IEEE, 2011,99(7): 1162–1182. [doi: 10.1109/JPROC.2011.2132790]
- [5] Saad W, Han Z, Hjørungnes A, Niyato D, Hossain E. Coalition formation games for distributed cooperation among roadside units in vehicular networks. IEEE Journal on Selected Areas in Communications, 2011,29(1):48–60. [doi: 10.1109/JSAC.2011.110106]
- [6] La C, Michiardi P, Casetti C, Chiasserini C, Fiore M. Content replication in mobile networks. IEEE Journal of Selected Areas in Communications, 2012,30(9):1762–1770. [doi: 10.1109/JSAC.2012.121021]
- [7] Vahdat A, Becker D. Epidemic routing for partially connected adhoc networks. Technical Report, CS-200006, Durham: Duke University, 2000.
- [8] Spyropoulos T, Psounis K, Raghavendra C. Spray and wait: An efficient routing scheme for intermittently connected mobile networks. In: Proc. of the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN). Philadelphia: ACM Press, 2005. [doi: 10.1145/1080139.1080143]
- [9] Jaho E, Koukoutsidis I, Stavrakakis I, Jaho I. Cooperative content replication in networks with autonomous nodes. Computer Communications, 2012,35(5):637–647. [doi: 10.1016/j.comcom.2011.12.006]
- [10] Cavaglione L, Cervellera C. An optimized content replication and distribution framework for vehicular networks. Journal of Intelligent Transportation Systems, 2011,15(4):179–192. [doi: 10.1080/15472450.2011.620474]
- [11] Kapadia S, Krishnamachari B, Ghandeharizadeh S. Static replication strategies for content availability in vehicular ad-hoc networks. Mobile Network Applications, 2009,14(5):590–610. [doi: 10.1007/s11036-008-0120-y]
- [12] Kumar N, Kim J. Probabilistic trust aware data replica placement strategy for online video streaming applications in vehicular delay tolerant networks. Mathematical and Computer Modelling, 2013,58(1):3–14. [doi: 10.1016/j.mcm.2012.07.021]
- [13] Zhuo X, Li Q, Gao W, Cao G, Dai Y. Contact duration aware data replication in delay tolerant networks. In: Proc. of the IEEE Int'l Conf. on Network Protocols (ICNP). Vancouver: IEEE Press, 2011. 236–245. [doi: 10.1109/ICNP.2011.6089057]
- [14] Wu Y, Zhu Y, Zhu H, Li B. CCR: Capacity-Constrained replication for data delivery in vehicular networks. In: Proc. of the IEEE Int'l Conf. on Computer Communications (INFOCOM). Turin: IEEE Press, 2013. 2580–2588. [doi: 10.1109/INFOCOM.2013.6567065]
- [15] Rezgui J, Cherkaoui S. About deterministic and non-deterministic vehicular communications over DSRC/802.11p. Wireless Communications & Mobile Computing, 2014,14(15):1435–1449. [doi: 10.1002/wcm.2270]
- [16] Keranen A, Ott J, Karkkainen T. The ONE simulator for DTN protocol evaluation. In: Proc. of the Int'l Conf. on Simulation Tools and Techniques (SIMUTools). Rome: ACM Press, 2009. [doi: 10.4108/ICST.SIMUTOOLS2009.5674]
- [17] Harri J, Filali F, Bonnet C. Mobility models for vehicular ad hoc networks: A survey and taxonomy. IEEE Communications Surveys & Tutorials, 2009,11(4):19–41. [doi: 10.1109/SURV.2009.090403]
- [18] Malandrino F, Casetti C, Chiasserini C, Fiore M. Content download in vehicular networks in presence of noisy mobility prediction. IEEE Trans. on Mobile Computing, 2014,13(5):1007–1021. [doi: 10.1109/TMC.2013.128]
- [19] Trullols-Cruces O, Fiore M, Barcelo-Ordinas J. Cooperative download in vehicular environments. IEEE Trans. on Mobile Computing, 2012,11(4):663–678. [doi: 10.1109/TMC.2011.100]



唐晓岚(1987—),女,河北邢台人,博士,讲师,CCF 会员,主要研究领域为车载网络,无线传感器网络,城市计算.



陈文龙(1976—),男,博士,副教授,CCF 会员,主要研究领域为互联网体系结构,物联网,高性能路由系统.



洪东惠(1993—),女,硕士生,主要研究领域为车辆自组织网络.



蒲菊华(1976—),女,博士,副教授,CCF 会员,主要研究领域为车辆自组织网络,数据挖掘,智慧城市,城市计算.