

基于车联网和移动边缘计算的时延可容忍数据传输

李萌^{1,2}, 司鹏搏^{1,2}, 孙恩昌^{1,2}, 张延华^{1,2}

(1. 北京工业大学信息学部, 北京 100124; 2. 北京工业大学先进信息网络北京实验室, 北京 100124)

摘要: 以物联网和车联网为代表的智慧城市的快速发展,使网络中的数据传输与数据计算面临巨大挑战,网络资源的分配也越来越受到广泛关注,为此提出了一种基于移动边缘计算的新型网络架构,通过整合物联网与车联网,用以传输时延可容忍数据及处理数据计算任务. 由于在同一网络架构下,需要融合多种网络标准和协议,基于可编程控制原理的软件定义网络技术被应用于所提网络架构中. 此外,时延可容忍数据在软件定义的车联网中的传输与计算节点选择过程可建模为部分可观测马尔科夫决策过程,从而优化并获得最小化系统开销,包括最小网络开销和最短数据计算处理时间. 仿真结果表明,与已有方案相比,所提方法可以有效地降低系统开销,缩短数据计算执行时间,提升数据计算效率,且在传输时延允许条件下,保证时延可容忍数据的传输到达率.

关键词: 车联网; 物联网; 时延可容忍数据; 移动边缘计算; 软件定义网络

中图分类号: TN 929 **文献标志码:** A **文章编号:** 0254 - 0037(2018)04 - 0529 - 09
doi: 10. 11936/bjtxb2017070032

Delay-tolerant Data Traffic Based on Connected Vehicle Network and Mobile Edge Computing

LI Meng^{1,2}, SI Pengbo^{1,2}, SUN Enchang^{1,2}, ZHANG Yanhua^{1,2}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
2. Beijing Laboratory of Advanced Information Networks, Beijing University of Technology, Beijing 100124, China)

Abstract: With the explosion in the number of Internet of things (IoT) and connected vehicle networks in smart city, the challenges to meet the demands from both data traffic delivery and data computing are increasingly prominent, and the allocation of network resources has attracted great attention. A novel network architecture based on mobile edge computing (MEC) was proposed in this paper to incorporate connected vehicle networks and IoT networks to transmit the delay-tolerant data and execute the computing tasks. In order to integrate diverse and complex standards and protocols in the same network, the programmable control principle originated from software-define networking (SDN) paradigm was introduced. Moreover, the process of delay-tolerant data transmission and computing node selection in software-defined vehicle network was formulated as a partially observable Markov decision process (POMDP) to minimize the system cost, which consists of both the network overhead and execution time of computing tasks. Simulation results show that the system cost can be decreased efficiently compared with the existing schemes, the processing time of computing tasks can be shorten and the computing efficiency can be improved. Furthermore, the arrival rate of delay-tolerant data can also be ensured

收稿日期: 2017-07-20
基金项目: 国家自然科学基金资助项目(61372089, 61571021, 61671029)
作者简介: 李萌(1988—), 男, 博士研究生, 主要从事无线网络资源分配、软件定义网络、物联网方面的研究, E-mail: limeng0720@emails.bjut.edu.cn
通信作者: 张延华(1960—), 男, 教授, 博士生导师, 主要从事信号处理方面的研究, E-mail: zhangyh@bjut.edu.cn

within the delay requirements.

Key words: connected vehicle network; Internet of things (IoT); delay-tolerant data; mobile edge computing (MEC); software-defined networking (SDN)

随着城市化进程的快速推进,智慧城市的建设成为大势所趋。作为智慧城市建设的的重要组成部分,物联网(Internet of things, IoT)的研究和应用受到了学术界和工业界的广泛关注^[1]。预计到2020年,用于IoT的设备或传感器将达到2 000亿台^[2]。因此,IoT的发展将导致数据传输的井喷式增长,同时,网络负载也势必随之提升。

此外,对于网络中部署的大量IoT设备(如智能门禁、智能检测、可穿戴设备等),其所包含的数据计算也将成为不可忽视的重要问题之一。然而,在IoT本地设备上进行复杂多样的数据运算将不可避免地导致设备负载提升、端到端计算数据时延及系统能效性降低等问题^[3]。

近些年,许多文献关注并研究改进IoT网络中的数据传输和数据计算问题,并取得了一定的效果。文献[3]考虑了将云计算技术引入到IoT网络中,通过构造基于IoT的云网络数学模型,实现数据传输和数据计算的网路资源分配。文献[4]为IoT网络中小数据包传输引入了2种新的无线通信接入协议,从而获得更高的数据传输吞吐量和更低的时延性能。文献[5]提出了一种在窄带(narrow band, NB)IoT系统中的小数据传输方案,它可使处于空闲状态的设备在没有网路资源控制连接建立过程的情况下,传输分组的小数据。文献[6]建立了一种基于IoT和云计算的联合策略,通过网路中的计算资源共享,获得计算数据执行的最大化。

然而,尽管针对IoT网络的数据传输和数据计算做了大量的研究工作,但当前研究仍存在2个容易被忽视的问题。一方面,当前网路的数据传输可分为时延敏感数据传输和时延可容忍数据传输。对于时延可容忍的数据传输,一定程度的传输时延增加并不严重影响此类数据的传输性能。然而,时延敏感数据和时延可容忍数据在传输过程中若不区分,将不可避免地造成网路资源负载增加和资源浪费^[7]。另一方面,传统的数据计算一般采用云计算模式,然而,数据存储与云计算服务器的部署距离通常很远,频繁的数据传输也会造成网路负载过重、网路资源不必要浪费等^[8]。

与此同时,智慧城市建设中的车联网也日益成为网路的重要组成部分之一。数以百万计的路侧单

元(roadside units, RSUs)以及装备有嵌入式设备的车辆可形成一个集数据通信、传输和计算的车联网。然而,车联网目前还仅仅关注于其本身的数据传输,通常仍具有空闲的网路资源,因此,该网路利用率较低,其中仍有许多空闲的网路资源可以被用于数据传输^[9]。文献[9]提出了一种整合车联网与数据网路的新型网路架构,使用车联网空闲资源来调度传输数据。此外,另一项关键技术——移动边缘计算(mobile edge computing, MEC)也逐渐成为无线网路数据计算的核心技术之一。MEC服务器的部署更靠近于终端用户,其可允许当前存储设备将计算任务卸载到MEC服务器上并执行该计算任务,从而更经济、有效地解决数据计算问题^[10]。

本文提出一种基于车联网和移动边缘计算时延可容忍的网路架构和数据传输优化方案,从而缓解网路数据传输拥塞,优化网路数据计算资源分配。由于该方案融合了多种网路标准和协议,文中通过软件定义网路(software-defined networking, SDN)进行网路资源的全局动态分配。此外,时延可容忍数据可选择通过车联网或IoT网路进行传输,数据计算任务可选择本地车辆或MEC服务器执行,而传输网路和计算节点的选择过程将被建模为部分可观测马尔科夫决策过程(partially observable Markov decision process, POMDP),根据网路每一时刻的状态以及观测的历史状态信息,做出网路传输和计算节点选择的最优决策,最终实现最小化系统开销(网路开销和计算任务执行时间)的优化目标。

1 系统模型

1.1 网路模型

本文所提的基于MEC的网路架构模型如图1所示。其中,包含了车联网与IoT网路,并随机分布了装备有嵌入式设备的车辆、RSU设备、IoT设施以及MEC服务器等。在每一时隙间隔,它们均处于可工作状态^[9]。当时延可容忍数据所在的当前存储需要数据传输或需要数据计算请求时,将会依据所提算法做出网路传输和计算节点选择的最优决策。当网路选择和计算服务器选择确定后,相应的设备和计算服务器将立即启动并做出相应的任务执行命令。该网路架构模型主要包含4个区域,具体描述

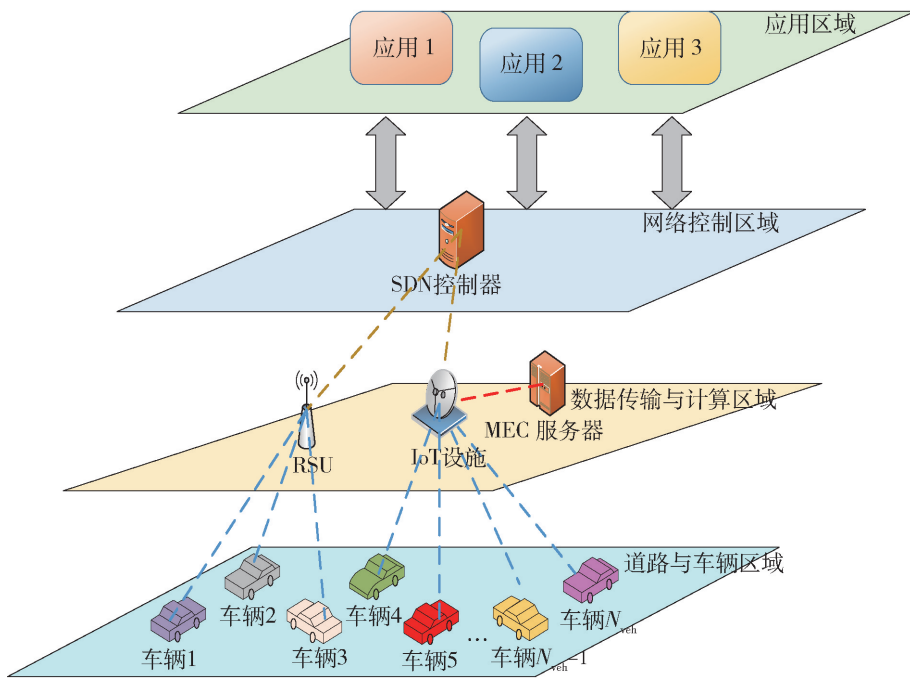


图1 智慧城市中软件定义的车联网与 MEC 网络模型

Fig. 1 Network model of software-defined vehicle networks with MEC in smart city

如下。

1.1.1 道路与车辆区域

为了模型简化,本文考虑的道路和车辆移动方向仅限于南-北和东-西直线方向。在道路与车辆区域内,假设有 N_{veh} 辆车随机分布在该区域内。同时,假设 M_{road} 表示道路的长度, M_{range} 表示车辆与 RSU 的通信覆盖范围。道路中,全部的路口数量表示为 N_{int} 。

1.1.2 数据传输与计算区域

在数据传输与计算区域中,随机分布了 N_{RSU} 个 RSU 和 N_{IoT} 个 IoT 设施(例如 NB-IoT 设施或 LoRa 模块)。同时,所有 IoT 设施边缘均部署了 MEC 服务器。此外,数据传输的时间考虑为等尺寸时隙,假设整个时间周期分为 K 个周期,每个时间点表示为 $t_0, t_1, \dots, t_k, \dots, t_{K-1}$, 每个时隙可表示为 $\delta t_k = t_k - t_{k-1}$ 。对于 IoT 网络中的最大允许传输时延表示为 T_{IoT} ,且满足 $\delta t_k \geq T_{IoT}$ 。

对于时延可容忍数据传输,在所提方案中考虑采用瑞利信道模型,假设车辆与 RSU 和 IoT 设施间可获得的数据传输速率分别表示为 H_{veh} 和 $H_{IoT}^{[11]}$ 。在网络架构中,获得的传输速率与信道状态信息有关(channel state information, CSI),根据瑞利信道分布特点,随机信道响应的概率密度函数表示为 $p_\theta(\theta) = (2\theta/\Omega) e^{-\theta^2/\Omega}$, 其中 $\theta \geq 0$ 且 $\Omega = E(\theta^2)$ 。

1.1.3 网络控制区域

在所提网络架构的网络控制区域,部署了基于 SDN 原理的控制器,用于灵活、可编程以及集中管理网络和计算资源^[12]。此外,SDN 控制器可以整合多种网络协议和标准,例如,从网络全局视野整合车联网与 IoT 网络。通过 SDN 控制器执行优化算法,最优传输网络决策以及最优计算服务器将会被确定,从而降低网络系统开销。

1.1.4 应用区域

应用区域包含了多项网络应用。网络设计的目标必须符合并满足这些网络应用的服务质量(quality of service, QoS)需求。

1.2 车辆移动模型

对于车辆的移动, $\gamma(k)$ 为载有时延可容忍数据的车辆移动方向的集合。在本文中,由于只考虑了道路或车辆(或所载数据)沿南-北、东-西方向的移动,因此, $\gamma(k)$ 包含了 5 种可能的情况:北向(N)、南向(S)、东向(E)、西向(W)以及静止 RSU 方向(F),即 $\gamma(k) \in \{N, S, E, W, F\}$ 。

在本文中,曼哈顿移动模型选取作为所提网络架构中车辆的移动模型^[13]。假设车辆和道路交口的密度分别为 ρ_{veh} 和 ρ_{int} , 车辆的移动速度为 v , 车辆随机移动的时间为 t_{move} 且满足均匀分布 $U(0, \frac{1}{\rho_{int}v})$ 。因此,车辆随机移动时间的概率密度函数可

以写为

$$f_{T_{\text{move}}}(t_{\text{move}}) = \begin{cases} \rho_{\text{int}} v, & t_{\text{move}} \in [0, \frac{1}{\rho_{\text{int}} v}] \\ 0, & \text{其他} \end{cases} \quad (1)$$

对于车辆到达道路路口,令 P_{wait} 表示车辆需要等待的概率,车辆在路口随机等待的时间表示为 t_{wait} ,假使随机等待时间满足均匀分布 $U(0, T_{\text{wait}})$,则概率密度函数可以表示为

$$f_{T_{\text{wait}}}(t_{\text{wait}}) = \begin{cases} \frac{1}{T_{\text{wait}}}, & t_{\text{move}} \in [0, T_{\text{wait}}] \\ 0, & \text{其他} \end{cases} \quad (2)$$

因此,车辆移动和停止的概率可以分别表示为

$$P_{\text{move}} = \frac{\frac{1}{\rho_{\text{int}} v}}{\frac{1}{\rho_{\text{int}} v} + \frac{T_{\text{wait}} P_{\text{wait}}}{2}} = \frac{2}{2 + T_{\text{wait}} P_{\text{wait}} \rho_{\text{int}} v} \quad (3)$$

$$P_{\text{stop}} = 1 - P_{\text{move}} = \frac{T_{\text{wait}} P_{\text{wait}} \rho_{\text{int}} v}{2 + T_{\text{wait}} P_{\text{wait}} \rho_{\text{int}} v} \quad (4)$$

式中: $\frac{1}{\rho_{\text{int}} v}$ 为车辆从一个路口移动到下一个路口的时间; $\frac{T_{\text{wait}}}{2}$ 为在路口处的平均等待时间; $\frac{1}{\rho_{\text{int}} v} + \frac{T_{\text{wait}} P_{\text{wait}}}{2}$ 为车辆移动和等待的全部时间。

此外,对于行驶在路上的车辆 A ,可以作为其邻居进行数据传输的情况分为 2 种:如果车辆 B 与车辆 A 行驶方向相同,且保持在相互通信范围内,则车辆 B 可作为车辆 A 的邻居,并建立直接通信链路;如果车辆 A 在路口处静止,其他等待在路口的车辆或路边 RSU 处于车辆 A 通信范围内,则这些车辆和 RSU 可作为车辆 A 的邻居并能与车辆 A 建立无线通信链路. 除这 2 种情况外,车辆 A 将无法选择邻居与其在车联网内建立无线通信链路。

1.3 计算模型

对于计算任务的执行,可以将计算任务分为本地计算服务器执行和 MEC 服务器执行. 也就是说,当时延可容忍数据需要进行计算处理时,需选择数据计算执行的服务器. 假设车辆 A 上需要执行的计算任务为 $I_{\text{vehA}} \Rightarrow (\alpha_{\text{vehA}}, \beta_{\text{vehA}})$. 式中: α_{vehA} 为需要计算数据的数据量大小; β_{vehA} 表示完成这些计算任务时, CPU 需要执行的循环次数^[14]. 本地计算和 MEC 的计算模型如下。

1.3.1 本地计算

装有嵌入式设备的车辆 A 处理计算数据的能力(如 CPU 循环处理能力)表示为 F_{vehA} ,则在第 k 个

时隙,其执行计算任务 I_{vehA} 的计算执行时间可表示为

$$t_{\text{vehA}}^l(k) = \frac{\beta_{\text{vehA}}(k)}{F_{\text{vehA}}} \quad (5)$$

1.3.2 MEC 计算

当计算任务在 MEC 服务器上执行时,计算任务将先卸载并通过无线通信链路传输到 MEC 服务器,计算任务从车辆 A 传输到 MEC 服务器的时间可表示为

$$t_{\text{vehA}, \text{off}}^m(k) = \frac{\alpha_{\text{vehA}}(k)}{H_{\text{IoT}}} \quad (6)$$

当数据卸载到 MEC 服务器后,数据计算将会执行,假设 MEC 服务器的计算数据处理能力为 F_{MEC} ,则数据计算的执行时间可表示为

$$t_{\text{vehA}, \text{comp}}^m(k) = \frac{\beta_{\text{vehA}}(k)}{F_{\text{MEC}}} \quad (7)$$

因此,数据计算在 MEC 服务器上的执行时间 $t_{\text{vehA}}^m(k)$ 可表示为

$$t_{\text{vehA}}^m(k) = t_{\text{vehA}, \text{off}}^m(k) + t_{\text{vehA}, \text{comp}}^m(k) \quad (8)$$

基于以上讨论,数据计算在本地服务器或 MEC 服务器执行的时间可以表示为

$$t_{\text{vehA}}(k) = \begin{cases} t_{\text{vehA}}^l(k), & \text{数据本地计算} \\ t_{\text{vehA}}^m(k), & \text{数据 MEC 服务器计算} \end{cases} \quad (9)$$

2 基于 POMDP 的网络与计算资源优化

2.1 POMDP 建模

POMDP 可以看作是一种特殊的马尔科夫模型,因为它强调状态转移过程中的“部分可观测”^[15]. 在软件定义的车联网中,由于时延可容忍数据的传输是否成功并不能准确获得,需要通过状态观测,因此,获得的最小系统开销的传输网络和计算节点选择问题可以建模为 POMDP 优化问题,通过包含动作和历史状态信息的状态观测信息,每个时隙内最优的动作决策将会被选择,从而使系统获得最低系统开销。

2.1.1 动作与策略

动作空间 \mathbf{A} 考虑为一个联合动作空间,其包含时延可容忍数据的网络传输选择和计算节点选择. 在 t_k 时刻,当前存储需要选择数据传输的网络和数据计算处理的服务器,因此, $a(k) \in \mathbf{A}$ 表示此时的混合动作,具体表示为

$$a(k) = [a_i(k), a_c(k)] \quad (10)$$

式中 $a_i(k)$ 和 $a_c(k)$ 分别为时延可容忍数据的网络

传输选择和数据计算节点选择. 具体来说, $a_i(k)$ 表示为

$$a_i(k) = \begin{cases} 0, & \text{数据保持在当前存储} \\ 1, & \text{数据通过 IoT 网络传输} \\ \dot{\gamma}, & \text{数据传输到邻居 } e' \end{cases} \quad (11)$$

式中: e' 为当前存储可选择传输的邻居, $\dot{\gamma} \in \{N, S, E, W, F\}$. 同时, $a_c(k)$ 可表示为

$$a_c(k) = \begin{cases} 0, & \text{数据在本地计算} \\ 1, & \text{数据在 MEC 服务器计算} \end{cases} \quad (12)$$

2.1.2 系统状态空间

系统状态空间集合定义为 \mathbf{S} . 在 t_k 时刻, 系统状态 $s(k) \in \mathbf{S}$ 包含了当前存储移动方向 $\gamma_{\text{cur}}(k)$ 、当前存储到目的地的距离 $d(k)$ 、数据成功传输指数 $\kappa(k)$. 因此, 系统状态 $s(k)$ 表示为

$$s(k) = [\gamma_{\text{cur}}(k), d(k), \kappa(k)] \quad (13)$$

此外, 数据成功传输指数 $\kappa(k)$ 可表示为

$$\kappa(k) = [\kappa_N(k), \kappa_S(k), \kappa_E(k), \kappa_W(k), \kappa_F(k)] \quad (14)$$

式中每个元素均为相应方向成功传输指数, 例如, $\kappa_N(k)$ 可以表示为

$$\kappa_N(k) = \begin{cases} 0, & \text{成功传输且 } \gamma_{e'}(k) = N \\ 1, & \text{其他} \end{cases} \quad (15)$$

在实际场景中, 当前存储可获得 $\gamma_{\text{cur}}(k)$ 和 $d(k)$ 的信息, 但是成功传输的信息 $\kappa(k)$ 很难直接、准确获得, 因此, 数据成功传输指数 $\kappa(k)$ 需要基于观测信息来准确获得.

2.1.3 系统观测空间

对于成功传输指数 $\kappa(k)$, 其观测空间集合定义为 \mathbf{O} . 在 t_k 时刻, 相应的观测状态表示为

$$o(k) = [o_N(k), o_S(k), o_E(k), o_W(k), o_F(k)] \quad (16)$$

式中每个元素均为相应方向的观测状态, 以 $o_N(k)$ 为例, 其可以表示为

$$o_N(k) = \begin{cases} 0, & \text{无邻居 } e' \text{ 且 } \gamma_{e'}(k) = N \\ 1, & \text{其他} \end{cases} \quad (17)$$

同时, 观测状态的概率定义为系统状态在动作 $a(k)$ 下, 从 $s(k) = i$ 转移到 $s(k+1) = j$ 时观测所产生的条件概率, 表示为

$$O_{ols}^a(i, j) = \frac{P_{\kappa|i, \gamma_{\text{cur}}}(i, j) P_{ol|\gamma_{\text{cur}}}(i, j)}{\sum_{\dot{\gamma} \in \mathbf{O}} P_{\kappa|i, \gamma_{\text{cur}}}(i, j) P_{ol|\gamma_{\text{cur}}}(i, j)} \quad (18)$$

通过引入该概率, 充分凸显了部分可观测马尔科夫决策过程中的“观测”特征, 从而解决在软件定义的车联网中, 由于时延可容忍数据的传输是否成

功并不能准确获得, 需要通过状态观测的问题. 在式(18)中, $P_{\kappa|i, \gamma_{\text{cur}}}(i, j)$ 和 $P_{ol|\gamma_{\text{cur}}}(i, j)$ 可分别表示为

$$P_{\kappa|i, \gamma_{\text{cur}}}(i, j) = \prod_{\dot{\gamma} \in \{N, E, S, W, F\}} P_{\kappa_{\dot{\gamma}}|i, \gamma_{\text{cur}}}(i, j) \quad (19)$$

$$P_{ol|\gamma_{\text{cur}}}(i, j) = \prod_{\dot{\gamma} \in \{N, E, S, W, F\}} P_{o_{\dot{\gamma}}|\gamma_{\text{cur}}}(i, j) \quad (20)$$

此时, 为计算式(19)(20), 有 2 种情况需要分别讨论, 具体情况如下:

当 $\gamma_{\text{cur}}(k) \neq F$ 时, $P_{o_{\dot{\gamma}}|\gamma_{\text{cur}}}(i, j)$ 可分别表示为

$$P_{o_F|\gamma_{\text{cur}}}(i, j) = \begin{cases} 1, & o_F(k) = 0 \\ 0, & o_F(k) = 1 \end{cases} \quad (21)$$

和

$$P_{o_{\dot{\gamma}}|\gamma_{\text{cur}}}(i, j) = \begin{cases} \left(1 - \frac{P_{\text{move}} M_{\text{range}}}{M_{\text{road}}}\right)^{\rho_{\text{veh}}}, & o_{\dot{\gamma}}(k) = 0 \\ 1 - \left(1 - \frac{P_{\text{move}} M_{\text{range}}}{M_{\text{road}}}\right)^{\rho_{\text{veh}}}, & o_{\dot{\gamma}}(k) = 1 \end{cases} \quad (22)$$

而 $P_{\kappa|i, \gamma_{\text{cur}}}(i, j)$ 可分别表示为

$$P_{\kappa_F|i, \gamma_{\text{cur}}}(i, j) = \begin{cases} 1, & \kappa_F(k) = 0 \\ 0, & \kappa_F(k) = 1 \end{cases} \quad (23)$$

和

$$P_{\kappa_{\dot{\gamma}}|i, \gamma_{\text{cur}}}(i, j) = \begin{cases} 1, & o_{\dot{\gamma}}(k) = 0 \text{ 且 } \kappa_{\dot{\gamma}}(k+1) = 0 \\ 0, & o_{\dot{\gamma}}(k) = 0 \text{ 且 } \kappa_{\dot{\gamma}}(k+1) = 1 \\ P_{\text{suc}}, & o_{\dot{\gamma}}(k) = 1 \text{ 且 } \kappa_{\dot{\gamma}}(k+1) = 0 \\ 1 - P_{\text{suc}}, & o_{\dot{\gamma}}(k) = 1 \text{ 且 } \kappa_{\dot{\gamma}}(k+1) = 1 \end{cases} \quad (24)$$

式中 P_{suc} 定义为数据成功从车辆 A 传输到其邻居的概率, 其可表示为

$$P_{\text{suc}} = \begin{cases} \frac{\rho_{\text{int}} v D_{\text{veh}A}}{H_*} - \frac{\rho_{\text{int}} v T_{\text{wait}}}{2}, & \frac{D_{\text{veh}A}}{H_*} \leq \min[t_{\text{more}} + t_{\text{wait}}, \delta t_k] \\ 0, & \frac{D_{\text{veh}A}}{H_*} > \min[t_{\text{more}} + t_{\text{wait}}, \delta t_k] \end{cases} \quad (25)$$

式中: $D_{\text{veh}A}$ 为传输数据包容量; $H_* \in \{H_{\text{veh}}, H_{\text{IoT}}\}$.

当 $\gamma_{\text{cur}}(k) = F$ 时, $P_{o_{\dot{\gamma}}|\gamma_{\text{cur}}}(i, j)$ 可分别表示为

$$P_{o_F|\gamma_{\text{cur}}}(i, j) = \begin{cases} 1 - P_{\text{int_RSU}}, & o_F(k) = 0 \\ P_{\text{int_RSU}}, & o_F(k) = 1 \end{cases} \quad (26)$$

和

$$P_{o_{\dot{\gamma}}|\gamma_{\text{cur}}}(i, j) = \begin{cases} \left(1 - \frac{P_{\text{move}} M_{\text{range}}}{M_{\text{road}}}\right)^{\rho_{\text{veh}}}, & o_{\dot{\gamma}}(k) = 0 \\ 1 - \left(1 - \frac{P_{\text{move}} M_{\text{range}}}{M_{\text{road}}}\right)^{\rho_{\text{veh}}}, & o_{\dot{\gamma}}(k) = 1 \end{cases} \quad (27)$$

而 $P_{\kappa_{lo}, \gamma_{cur}}(i, j)$ 可分别表示为

$$P_{\kappa_{\dot{\gamma}}, \gamma_{cur}}(i, j) = \begin{cases} 1, & \kappa_{o_{\dot{\gamma}}}(k+1) = 0 \\ 0, & \kappa_{o_{\dot{\gamma}}}(k+1) = 1 \end{cases} \quad (28)$$

2.1.4 状态转移

系统状态的一步转移概率定义为 $P_s^a(i, j)$, 其可表示为

$$P_s^a(i, j) = \sum_{o \in O} \left[P_{\kappa_{lo}, \gamma_{cur}}(i, j) P_{\gamma_{cur|a}}(i, j) \prod_{\dot{\gamma} \in \{N, S, W, E, F\}} P_{o_{\dot{\gamma}} | \gamma_{cur}}(i, j) \right] \quad (29)$$

$P_{\kappa_{lo}, \gamma_{cur}}(i, j)$ 和 $P_{o_{\dot{\gamma}} | \gamma_{cur}}(i, j)$ 在 2.1.3 小节已经给出了计算方法, $P_{\gamma_{cur|a}}(i, j)$ 则需根据不同的动作给定, 具体计算过程表示如下:

当 $a_t(k) \in \{0\}$ 时, $a_c(k) \in \{0, 1\}$, 在此条件下, 时延可容忍数据将会保持在当前存储, 计算任务可选择在本地或 MEC 服务器执行. 此时, 概率 $P_{\gamma_{cur|a}}(i, j)$ 可表示为

$$P_{\gamma_{cur|a}}(i, j) = \begin{cases} 1, & \gamma_{cur}(k) = F, \gamma_{cur}(k+1) = F \\ P(\dot{\gamma}_j | \dot{\gamma}_i), & \gamma_{cur}(k) = \dot{\gamma}_i, \gamma_{cur}(k+1) = \dot{\gamma}_j \\ 0, & \text{其他} \end{cases} \quad (30)$$

当 $a_t(k) \in \{1\}$ 时, $a_c(k) \in \{1\}$, 在此条件下, 时延可容忍数据将会通过 IoT 设施传输, 计算任务可选择在 MEC 服务器执行. 此时, 概率 $P_{\gamma_{cur|a}}(i, j)$ 可表示为

$$P_{\gamma_{cur|a}}(i, j) = \begin{cases} 0, & \gamma_{cur}(k+1) = \dot{\gamma}_i \\ 1, & \text{其他} \end{cases} \quad (31)$$

当 $a_t(k) \in \{N, E, S, W, F\}$ 时, $a_c(k) \in \{0, 1\}$, 在此条件下, 时延可容忍数据将会通过车联网传输, 计算任务可选择在本地或 MEC 服务器执行. 此时, 概率 $P_{\gamma_{cur|a}}(i, j)$ 可表示为

$$P_{\gamma_{cur|a}}(i, j) = \begin{cases} P_{suc} P(\dot{\gamma}_j | \dot{\gamma}_i), & \gamma_{cur}(k) = \dot{\gamma}_i, \gamma_{cur}(k+1) = \dot{\gamma}_j \\ 0, & \text{其他} \end{cases} \quad (32)$$

2.1.5 系统开销

在软件定义的车联网中, 当前存储可通过所提方法来选取最优决策, 从而获得最小系统开销, 即表示为最小网络开销和最小数据计算执行时间. 对于车辆 A , 其系统开销 $Q_{vehA}(k)$ 可以表示为

$$Q_{vehA}(k) = \begin{cases} \eta t_{vehA}(k) + \zeta C_{veh}(k), & a_t(k) \in \{0, \dot{\gamma}\} \\ \eta t_{vehA}(k) + \zeta C_{IoT}(k), & a_t(k) \in \{1\} \end{cases} \quad (33)$$

式中: η 和 ζ 为权重因子, 并满足 $0 \leq \eta, \zeta \leq 1$, 且 $\eta + \zeta = 1$; C_{veh} 和 C_{IoT} 分别为当前存储选择车联网和 IoT 网络时的网络开销.

对于系统在整个时间周期的全部期望开销可以表示为

$$Q = \arg \min_{\{a_t, a_c\}} \left[\sum_{k=0}^K \sum_{vehA=1}^{N_{veh}} \sigma^{K-k} Q_{vehA}(k) \right] \quad (34)$$

式中 σ 为折扣因子, 其满足 $0 \leq \sigma \leq 1$. 系统中的各用户在每个时间周期内, 均通过选择网络传输和计算服务器的最优选择决策, 实现式 (34) 的优化目标. 同时, 折扣因子 σ 用以保证所提优化算法具有收敛性, 确保可求出算法最优解.

2.2 优化问题求解

通过以上讨论, 基于 POMDP 的最优策略建模已给定, 对于函数求解问题将借助于动态的值迭代和基于贝叶斯的信息状态更新^[16]. 给定值函数 $\varphi_k(\pi(k))$, 其定义为在时隙 $\delta t(k)$ 时最小系统开销, 具体表示为

$$\varphi_k(\pi(k)) = \min_{a(k) \in A} \left\{ \sum_{i \in S} \sum_{j \in S} \pi_i^k P_s^a(i, j) \sum_{o(k) \in S} O_{ols}^a(i, j) [Q_{vehA}(k) + \varphi_{k+1}(\pi(k+1))] \right\} \quad (35)$$

式中 $\pi(k)$ 为信息空间, 可表示为

$$\pi(k) = \{\pi_1^k, \pi_2^k, \dots, \pi_i^k, \pi_j^k, \dots, \pi_s^k\} \quad (36)$$

对于 π_i^k , 可通过贝叶斯更新规则获得, 具体表示为

$$\pi_j^{k+1} = \frac{\sum_i \pi_i^k P_s^a(i, j) O_{ols}^a(i, j)}{\sum_i \sum_j \pi_i^k P_s^a(i, j) O_{ols}^a(i, j)} \quad (37)$$

通过贝叶斯更新迭代计算, 求解 POMDP 中的信息状态概率. 根据式 (37), 每一时刻的状态转移概率和观测概率均能够映射在信息状态中, 通过迭代计算, 可获得在每一时刻的基于状态转移和状态观测信息的信息状态概率, 为优化决策的选取提供依据.

对于所提的 POMDP 算法, 计算复杂度依赖于 POMDP 建模的动作空间和状态转移空间的矩阵维数. 在本文所提的网络模型中, 状态转移空间和动作空间所构成的矩阵维数分别为 $5 \times 5 \times D$ 和 7×2 , 其中 D 为状态 $d(k)$ 的离散等级数量. 在实际过程中, 优化选择过程需要执行以上空间维数的计算和

函数迭代. 同时, 该计算复杂度也可能会随着动作空间和状态维数空间的增加而增加. 与已有网络模型不同, 本文所提系统模型中由于引入了 SDN 控制器, 这些复杂的计算过程将在线下执行, 大幅降低系统的实时计算量. 当给定所有参数的值时, 系统最小开销的最优动作决策将在线下运行计算并确定. 同时, 如果参数发生变化, 值函数也会在线下重新计算并动态更新^[17]. 也就是说, 一旦确定所有参数, 那么最优策略即可获得, 并将其存储在控制器中. 而在线上, 系统将不再重复执行复杂的计算过程, 当前存储只依据当前状态选择相应的最优动作^[18-20].

3 仿真结果与分析

本节将构建车联网和移动边缘计算架构的仿真环境, 通过 Matlab 软件仿真验证基于 POMDP 的最小网络开销和计算数据执行时间的优化算法的优越性. 结合参考文献[9][21]和[22]以及实际环境, 具体仿真参数设置如表 1 所示.

表 1 仿真参数设置

Table 1 Setting of simulation parameter

参数名称	参数值
网络包含车辆总数/辆	20 000
车辆移动速度/($\text{m}\cdot\text{s}^{-1}$)	8.33
相邻路口距离/m	500
路口密度/ m^{-1}	0.002
车辆路口等待概率	0.8
车辆路口等待时间/s	150
计算数据容量/MB	15
车载 CPU 性能/GHz	0.5
MEC 服务器性能/GHz	1
车联网数据传输开销	10
IoT 网络数据传输开销	$20d(k)$
权重因子	0.5

为公平对比所提方法优越性, 本文将所提方法与仅使用车联网且本地计算和仅使用 IoT 网络且 MEC 的已有方法进行对比验证.

如图 2 所示, 相比于仅使用车联网和仅使用 IoT 网络, 基于所提方法的系统开销明显降低. 主要原因是当前存储所做的最优决策是依据网络状态和观测状态, 因此, 当前存储将选择最优的网络进行数据传输和最优计算服务器处理数据计算. 当仅使用车

联网进行数据传输和本地执行数据计算任务时, 系统开销显著增高, 主要原因是本地执行数据计算任务将导致计算处理时间提升, 系统开销也随之提升. 而对于仅使用 IoT 网络和 MEC 进行数据传输与计算, 尽管该方案可保证短时间处理计算任务, 但网络传输的使用成本较高. 因此, 本文所提方法相比于其他 2 种方法, 可获得最小系统开销.

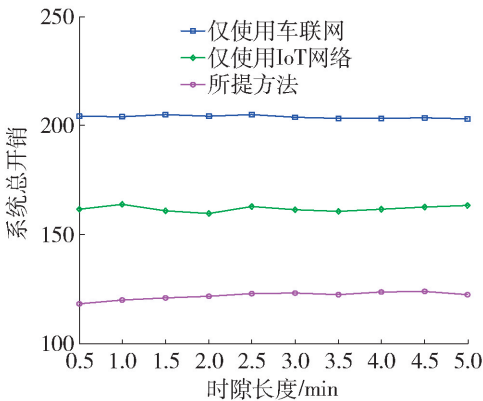


图 2 不同时隙长度条件下系统开销比较

Fig. 2 System cost comparison with different time slot lengths

图 3 验证了在不同 CPU 计算循环次数的条件下, 基于 3 种方案的系统开销对比. 相比于仅使用车联网和 IoT 网络, 所提方案的系统开销随着计算任务的 CPU 计算循环次数增加而缓慢提升. 主要原因是在所提方案下, 当前存储可根据网络条件动态选择计算服务器. 也就是说, 根据网络条件和状态转移过程, 通过决策选择最优的数据计算服务器. 由于所提方案考虑了本地计算与 MEC 服务器的动态选择, 因此, 可有效降低本地计算的系统开销, 提升数据计算处理的时效性.

图 4 验证了不同时间长度条件下, 数据传输在时延要求内的到达率. 由图 4 可以看出, 所提方法的数据传输到达率在各个时隙内均保持在 90% 以上, 与仅使用 IoT 网络条件下的数据到达率基本相同, 从而保证了数据传输的可靠性. 而对于仅使用车联网方案, 由于数据仅可在车辆或 RSU 之间传输, 因此, 当传输时间达到时延要求时, 未传到目的地的数据不得被丢弃, 从而不能保证数据准时到达目的地, 降低了数据的到达率.

综上所述, 相比于仅使用车联网且本地数据计算或仅使用 IoT 网络且 MEC 服务器执行数据计算, 所提方法在保证时延可容忍数据到达率的前提下, 能够有效降低网络使用开销和数据计算执行时间,

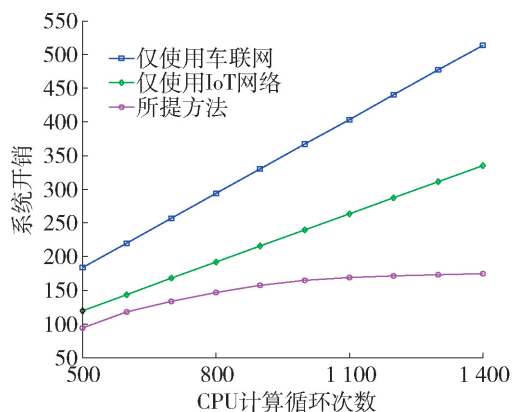


图3 不同CPU计算循环次数条件下系统开销比较

Fig. 3 System cost comparison with different number of computing cycles

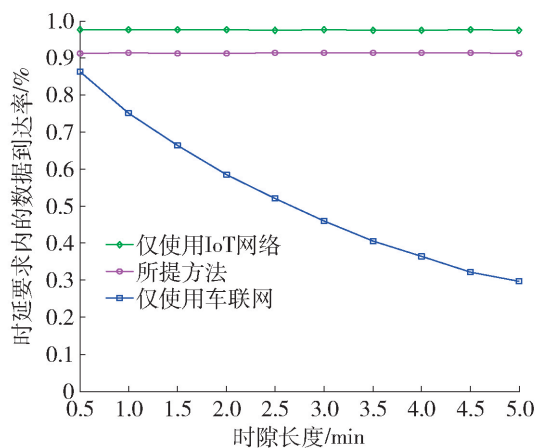


图4 不同时隙长度条件下数据到达率比较

Fig. 4 Data arrival rate comparison with different time slot lengths

从而更显著地降低系统全局开销,提升网络数据传输和数据计算的效率。

4 结论

1) 提出了一种新的网络架构,即在智慧城市的IoT网络中,考虑可容忍时延数据传输的特性,通过车联网的空闲网络资源进行数据传输。同时,在所提网络架构中引入了移动边缘计算,使数据计算任务可以选择在本地服务器或MEC服务器上执行。

2) 通过加入SDN控制器,使多种网络协议和标准在同一网络架构下得到整合,网络资源可以动态灵活调度与分配。

3) 数据传输网络与数据计算执行服务器的选择过程建模为POMDP优化策略,通过观测系统状态,使当前存储对数据传输和计算节点的选取做出最优决策,从而最小化系统开销。与已有方案相比,

所提方案可在时延允许范围内、最小系统开销条件下进行数据传输与数据计算处理。

4) 未来,结合本文所提的网络架构和优化模型,重点考虑智慧城市IoT网络与车联网相结合时的能源利用率问题,优化整合网络的系统能效性,进一步降低网络开销成本。

参考文献:

- [1] EJAZ W, NAEEM M, SHAHID A, et al. Efficient energy management for the internet of things in smart cities [J]. IEEE Communications Magazine, 2017, 55(1): 84-91.
- [2] AL-FUQAHA A, GUIZANI M, MOHAMMADI M, et al. Internet of things: a survey on enabling technologies, protocols, and applications [J]. IEEE Communications Surveys & Tutorials, 2015, 17(4): 2347-2376.
- [3] BARCELO M, CORREA A, LLORCA J, et al. IoT-cloud service optimization in next generation smart environments [J]. IEEE Journal on Selected Areas in Communications, 2016, 34(12): 4077-4090.
- [4] CENTENARO M, VANGELISTA L, SAUR S, et al. Comparison of collision-free and contention-based radio access protocols for the internet of things [J]. IEEE Transactions on Communication, 2017, 99: 1-15.
- [5] OH S-M, SHIN J. An efficient small data transmission scheme in the 3GPP NB-IoT system [J]. IEEE Communications Letters, 2017, 21(3): 660-663.
- [6] FARRIS I, MILITANO L, NITTI M, et al. Federated edge-assisted mobile clouds for service provisioning in heterogeneous IoT environments [C] // IEEE 2nd World Forum Internet of Things (WF-IoT). New York: IEEE, 2015: 591-596.
- [7] ZHANG Y, ZHANG Y, QIN S, et al. Delay-bounded priority-driven resource allocation for video transmission over multihop networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(7): 1184-1196.
- [8] CAI Y, YU F R, BU S. Cloud computing meets mobile wireless communications in next generation cellular networks [J]. IEEE Network, 2014, 28(6): 54-49.
- [9] SI P, HE Y, YAO H, et al. DaVe: offloading delay-tolerant data traffic to connected vehicle networks [J]. IEEE Transactions on Vehicular Technology, 2016, 65(6): 3941-3953.
- [10] CHIANG M, ZHANG T. Fog and IoT: an overview of research opportunities [J]. IEEE Internet of Things Journal, 2016, 3(6): 854-864.
- [11] KUMAR N, ZEADALLY S, RODRIGUES J. Vehicular

- delay-tolerant networks for smart grid data management using mobile edge computing [J]. IEEE Communications Magazine, 2016, 54(10): 60-66.
- [12] CHEN T, MATINMIKKO M, CHEN X, et al. Software defined mobile networks: concept, survey, and research directions [J]. IEEE Communications Magazine, 2015, 53(11): 126-133.
- [13] HARRI J, FILALI F, BONNET C. Mobility models for vehicular ad networks: a survey and taxonomy [J]. IEEE Communications & Surveys Tutorials, 2009, 11(4): 19-41.
- [14] CHEN X. Decentralized computation offloading game for mobile cloud computing [J]. IEEE Transactions Parallel and Distributed Systems, 2015, 26(4): 974-983.
- [15] ZHAO Q, TONG L, SWAMI A, et al. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework [J]. IEEE Journal on Selected Areas in Communications, 2007, 25(3): 589-600.
- [16] CHEN Y, ZHAO Q, SWAMI A. Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors [J]. IEEE Information Theory, 2008, 54(5): 2053-2071.
- [17] SI P, YU F R, JI H, et al. Distributed sender scheduling for multimedia transmission in wireless mobile peer-to-peer networks [J]. IEEE Transactions on Wireless Communications, 2009, 8(9): 4594-4603.
- [18] ZHU L, YU F R, NING B, et al. Handoff performance improvements in MIMO-enabled communication-based train control systems [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2): 582-593.
- [19] LI M, YU F R, SI P, et al. Energy-efficient M2M communications with mobile edge computing in virtualized cellular networks [C] // 2017 IEEE International Conference on Communications (ICC). New Yoirk: IEEE, 2017: 1-6.
- [20] LI M, YU F R, SI P, et al. Random access and virtual resource allocation in software-defined cellular networks with machine-to-machine (M2M) communications [J]. IEEE Transactions on Vehicular Technology, 2017, 66(7): 6399-6414.
- [21] BAI F, SADAGOPAN N, HELMY A. Important: a framework to systematically analyze the impact of mobilityon performance of routing protocols for Adhoc networks [C] // 2003 IEEE INFOCOM. New York: IEEE, 2003: 825-835.
- [22] PADMANABHAN S, STEPHEN R, MURTHY C, et al. Training-based antenna selection for PER minimization: a POMDP approach [J]. IEEE Transactions on Communications, 2015, 63(9): 3247-3260.

(责任编辑 吕小红)