第 39 卷 第 7 期
2016 年 7 月

合肥工业大学学报(自然科学版)
JOURNAL OF HEFEI UNIVERSITY OF TECHNOLOGY (Natural Science)

Vol. 39 No. 7
Jul. 2016

DOI:10.3969/j.issn.1003-5060.2016.07.007

一种基于多 Agent 强化学习的无线传感器网络多路径路由协议

乔 阳[1], 唐 昊[1], 程文娟[2], 江 琦[1], 马学森[2]

(1.合肥工业大学 电气与自动化工程学院,安徽 合肥 230009; 2.合肥工业大学 计算机与信息学院,安徽 合肥 230009)

摘 要:文章研究了无线传感器网络中存在的多条最短路径路由选择问题。将无线传感器网络看作多 Agent 系统,采用强化学习理论,提出了一种基于多 Agent 强化学习的无线传感器网络多路径路由协议 MRL-MPRP(Multi-agent Reinforcement Learning based Multiple-path Routing Protocol)。该协议综合考虑了所要发送数据的优先级、节点间的链路质量以及节点数据缓冲队列的拥堵情况,为不同优先级的数据选择出当前网络状况下最优的路径进行数据的传输。仿真结果表明了该协议在降低网络平均端—端延时、提升数据包成功投递率方面的有效性。

关键词:无线传感器网络;多路径路由协议;多 Agent 系统;强化学习

中图分类号:TP181;TP212.9 **文献标识码**:A **文章编号**:1003-5060(2016)07-0896-04

A multiple-path routing protocol in wireless sensor network based on multi-agent reinforcement learning

QIAO Yang[1], TANG Hao[1], CHENG Wenjuan[2], JIANG Qi[1], MA Xuesen[2]

(1. School of Electric Engineering and Automation, Hefei University of Technology, Hefei 230009, China; 2. School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract: In this paper, the optimal route selection problem in the case of several shortest paths with the same hops in wireless sensor networks is considered. A multi-agent reinforcement learning based multiple-path routing protocol(MRL-MPRP) is proposed by regarding wireless sensor network as a multi-agent system. In MRL-MPRP, the sensor node takes the priority of the transmitting data, link quality and congestion of different neighbors into consideration so as to select an optimal route for sending different types of data. The simulation results show that the proposed protocol effectively reduces the end-to-end delay and increases the packet delivery ratio.

Key words: wireless sensor network; multiple-path routing protocol; multi-agent system; reinforcement learning

无线传感器网络是一种面向任务的网络,在环境监测、电力系统监控、地震监测等应用中,节点所采集到的数据重要性各不相同,一些异常数据(如温度过高)通常反映了被监控系统所出现的反常现象,需要确保能够及时、准确地传送至控制中心,对延时、数据投递成功率等网络性能都提出了较高的要求[1-2]。在传统的多路径路由协议中,一些算法通过在最短路径上传输优先级较高的异

收稿日期:2015-03-13;修回日期:2015-04-30

基金项目:国家自然科学基金资助项目(61174186;61374158;71231004;51274078);教育部新世纪优秀人才计划资助项目(NCET-11-0626)和高等学校博士学科点专项科研基金资助项目(20130111110007)

作者简介:乔 阳(1990—),男,安徽淮北人,合肥工业大学硕士生;

唐 昊(1972—),男,安徽庐江人,博士,合肥工业大学教授,博士生导师;

程文娟(1970—),女,安徽怀宁人,博士,合肥工业大学教授,硕士生导师.

常数据，而在其他非最短路径上传输优先级较低的常规数据从而满足不同数据对传输性能的要求[3-4]。文献[5-7]也提出了几种多路径路由协议以保证网络在传输数据时的服务质量。

然而，由于无线传感器网络中的节点是通过自组织的方式相互连接的，网络的拓扑结构具有高度动态性，对某些节点的数据传输而言，网络中可能同时存在多条跳数相同的最短路径。受节点部署位置、传输干扰以及事件触发频率等环境因素的影响，这些路径的链路状况和拥堵情况具有差异性，其在传输数据时的服务质量也各不相同。因此，需要根据网络的实时动态从多条最短路径中为不同类型的数据寻找出相应的最优路径进行数据的传输，尤其是那些优先级较高的实时数据。

针对上述问题，本文提出了一种基于多 Agent 强化学习的无线传感器网络多路径路由协议 MRL-MPRP（Multi-agent Reinforcement Learning based Multiple-path Routing Protocol）。将无线传感器网络看作 1 个多 Agent 系统，网络中的每个节点都是 1 个具有独立学习能力的智能体（Agent），节点在路由选择时，考虑所发送数据的优先级、与各邻居节点之间的链路质量以及各邻居节点数据缓冲队列长度等网络实时信息。同时，采用强化学习理论，将高优先级数据的路由选择过程建模为 Markov 决策过程，并引入基于分布式值函数的 Q 学习算法求解最优路径，通过节点间的局部信息交互实现全局最优。

1　网络模型

1.1　节点模型

将节点采集到的数据和接收到的数据统称为到达节点的数据，分为高、低 2 个优先级。同时，假设节点自身采集的数据是随机到达的，高、低优先级数据的到达间隔分别服从参数为 λ_1 和 λ_2 的指数分布，数据包大小均为 C。假设每个节点均配备一个缓冲队列，用于存放到达节点的数据，队列的容量为 M，数据放入缓冲队列后等待传输，若数据到达时缓冲队列已满，则丢弃相应的数据包。队列中的数据按照“先进先出”原则发送，即发送顺序与优先级无关。

1.2　信道模型

在数据传输时，对于衰落信道而言，信号通过无线信道后，其幅值是随机的，假设节点接收到的瞬时信噪比（signal-to-noise ratio，SNR）A 呈指数分布。令 $0=A_0<A_1<A_2<\cdots<A_H=\infty$ 表示接收的 SNR 的门限值，若接收的 SNR 落在区间 $[A_h,A_{h+1})$，$h=0,1,\cdots,H-1$，则称信道处于 L_h 状态，即发送节点和接收节点之间的链路质量为 L_h 状态[8]。

2　MRL-MPRP 协议

2.1　协议工作机制

在上述网络模型下，考虑一个有 N 个节点的无线传感器网络，假设节点的路由表中已保存到终端节点的所有最短路径。当节点 i 的数据缓冲队列不为空，即节点 i 有通信需求时则启动路由发现机制，具体过程如下：

若所要发送的数据优先级为高，则向其路由表中保存的所有下一跳邻居节点（在文中均指最短路径上的下一跳节点）组播路由请求包 RREQ。邻居节点收到 RREQ 消息后，会根据接收信号强度，计算自身与发送节点之间的链路质量，同时读取自身的缓冲队列长度，然后向节点 i 发送路由应答信息包 RREP，并在 RREP 消息中附上上述链路质量和缓冲队列长度。发送节点 i 在收到所有邻居节点回送的 RREQ 消息后，根据自身与各邻居节点之间的链路质量、邻居节点的队列长度做出决策，从邻居节点中选择符合要求的 1 个作为下一跳节点，并发送数据。若发送数据的优先级为低，节点从路由表中随机挑选 1 个邻居作为下一跳节点。

邻居节点 j 在收到节点 i 发送的数据后，若缓冲队列不为满，则将数据包放入队列，并向节点 i 发送确认字符（ACK）消息；若队列已满，则丢弃相应的数据包，并向节点 i 发送否定确认（NACK）消息；若节点 i 在一段时间后既未收到 ACK 也未收到 NACK 消息，则认为数据包在传输过程中丢失，重新发送数据包直至收到确认信息（ACK、NACK）或者达到最大重传次数为止，此时认为当前数据包的发送已完成。随后，节点 i 会检查自身的缓冲队列，若队列中有数据包，则进入下一数据包的路由选择过程；否则，节点 i 一直等待直到下一个数据包到达。

2.2　数学建模

对高优先级数据的路由选择过程，采用多 Agent 强化学习理论来实现，将其建立为一个马尔可夫决策过程（Markov decision process，MDP）模型[9]。并引入基于分布式值函数的 Q 学习算法来求解最优路径，通过仅与通信范围内的邻居节点进行局部信息交互实现全局最优[10]。节点 i

第 n 次高优先级数据路由选择的 MDP 过程可定义如下。

(1) 状态。节点 i 自身的状态 s_i^n 定义为:

$$s_i^n = \{(l_{ij}^n, q_j^n) \mid j \in I_i\} \tag{1}$$

其中,l_{ij}^n 为节点 i 与邻居节点 j 之间的链路质量,$l_{ij}^n \in \{L_0, L_1, \cdots, L_{H-1}\}$;$q_j^n$ 为邻居节点 j 的队列长度,$q_j^n \in \{0, 1, 2, \cdots, M\}$;$I_i$ 为邻居节点的集合。并记 s_i^n 所有可能取值的集合为 S_i。

(2) 行动。节点 i 当前采取的行动 $a_i^n(s_i^n) \in D_i$,且仅与自身当前的状态有关,D_i 定义为:

$$D_i = \{\mathrm{ch}_j, j \in I_i\} \tag{2}$$

其中,ch_j 表示节点 i 选择节点 j 为下一跳节点,并发送数据。

(3) 代价函数。若被选节点 j 成功收到节点 i 发送的数据包,则会在回送的 ACK 消息中附上数据到达节点 j 的时刻。节点 i 根据 $t_{\mathrm{tx}} = t_{\mathrm{arrival}} - t_{\mathrm{leaving}}$ 计算该高优先级数据包传输所经历的延时,其中 t_{leaving} 为数据包离开节点 i 的时刻。节点 i 在一次高优先级数据传输中的立即代价定义如下:

$$r_i^n(s_i^n, a_i^n(s_i^n)) = \begin{cases} K_1 \dfrac{t_{\mathrm{tx}}}{t_{\mathrm{avg}}} + K_2 \dfrac{q_j^n}{M}, & \text{收到 ACK;} \\ K_{\mathrm{drop}}, & \text{收到 NACK;} \\ K_{\mathrm{lost}}, & \text{数据包丢失} \end{cases} \tag{3}$$

其中,t_{avg} 为数据包传输的平均时间,$t_{\mathrm{avg}} = 1/R_{\mathrm{tx}}$,$R_{\mathrm{tx}}$ 为传输速率;q_j^n 为被选节点 j 的队列长度;K_1 为传输延时在代价中所占比重;K_2 为队列长度在代价中所占比重;K_{drop} 为收到 NACK 所获得的代价;K_{lost} 为数据包丢失获得的代价。

(4) Q 值更新。节点 i 在获得立即代价后,使用基于分布式值函数的 Q 学习算法更新当前状态-行动对的 Q 值,更新公式为:

$$\begin{aligned} Q_i^{n+1}(s_i^n, a_i^n(s_i^n)) = {} & (1-\alpha) Q_i^n(s_i^n, a_i^n(s_i^n)) + \alpha\big[r_i^n(s_i^n, a_i^n(s_i^n)) + \\ & \gamma\omega(i,j) \min_{s_j^n \in S_j, a_j^n \in D_j} Q_j^n(s_j^n, a_j^n(s_j^n)) + \\ & \gamma \sum_{i' \in I_i, i' \neq j} \omega(i,i') \min_{s_{i'}^n \in S_{i'}, a_{i'}^n \in D_{i'}} Q_{i'}^n(s_{i'}^n, a_{i'}^n(s_{i'}^n))\big] \end{aligned} \tag{4}$$

其中,α 为学习率;γ 为折扣因子;$\omega(i,j)$ 为节点 i 从被选节点 j 获得的长远代价的权重;$\omega(i,i')$ 为节点 i 从其他邻居节点获得的长远代价的权重。

3 仿真实验及结果分析

本文以 OMNET++4.0 作为实验仿真平台,验证所提出的 MRL-MPRP 协议对网络性能的影响。假设 45 个节点随机分布在 200 m×200 m 的区域内,网关节点处于区域的中心,节点的通信距离为 50 m。其他主要仿真参数设置如下:仿真时间 t=1 600 s,数据包大小 C=100 bit,队列容量 M=4,链路状态数 H=3,传输速率 R_{tx}=100 kb/s,传输功率 P_{tx}=4 mW,高、低优先级数据包到达间隔 λ_1、λ_2 均为 15 ms。

高、低 2 个优先级数据包的平均传输端—端延时如图 1 所示。由图 1 可以看出,随着仿真的运行,高优先级数据包的平均传输端—端延时呈下降趋势,至仿真结束时,其延时比仿真初始阶段下降约 15%,而低优先级数据包的平均端—端延时则在仿真过程中无明显变化,且其波动相对较大。这说明所提出的 MRL-MPRP 协议可以有效减少高优先级数据的传输延时。

网络中的节点在一跳传输中成功发送 1 个高、低优先级数据包所需重传的平均次数如图 2 所示。

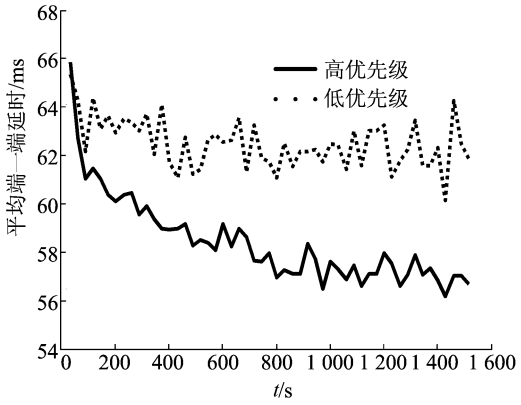

图 1 高、低优先级数据包平均传输端—端延时

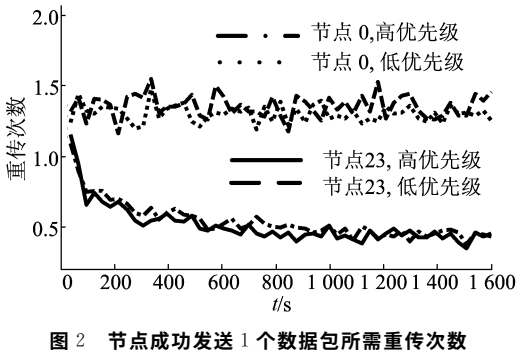

图 2 节点成功发送 1 个数据包所需重传次数

从图 2 可以看出,随着仿真的进行,所选节点成功发送高优先级数据包所需重传的次数逐渐降低,且最终的值明显低于低优先级数据。这是因为 MRL-MPRP 协议在发送高优先级数据优先级时,考虑节点之间的链路质量,在链路质量较好的路径中传输数据可以有效提高数据包发送成功的

概率，因此其失败重传的次数会降低。

高、低 2 个优先级数据包的投递成功率曲线如图 3 所示。由图 3 可以看出，在仿真过程中，高优先级数据包的投递成功率逐渐增大，而低优先级数据的投递成功率则变化不大，统计意义上来说，略有提高；且高优先级数据的投递成功率明显高于低优先级数据，最终约高出 10.4%。这说明所提出的 MRL-MPRP 协议可以有效提高高优先级数据包的投递成功率。

高优先级数据包的投递成功率与数据包产生间隔时间之间的关系如图 4 所示。

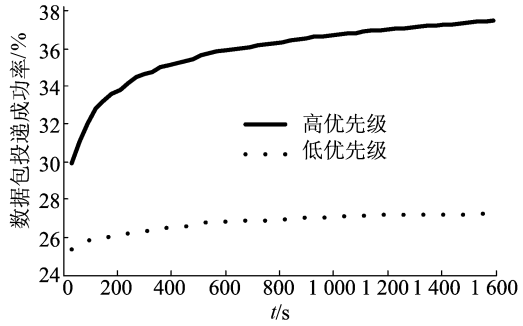

图 3　高、低优先级数据包投递成功率对比曲线

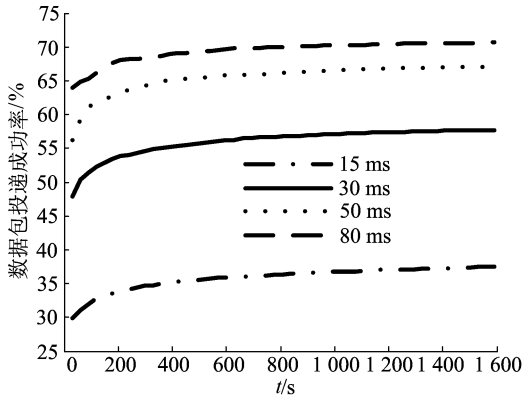

图 4　不同到达间隔时间下高优先级数据包投递成功率

从图 4 可以看出，在每个间隔时间下，仿真结束时的投递成功率都比仿真开始时有一定程度的提升，最高可提升 12% 左右（间隔 50 ms 时）。这说明所提出的 MRL-MPRP 协议对不同流量强度的网络均有不同程度的适应性。

4　结　论

本文研究了无线传感器网络的多条最短路径寻优问题，将无线传感器网络作为一个多 Agent 系统，提出了一种基于多 Agent 强化学习的多路径路由协议 MRL-MPRP。若发送的是高优先级数据，协议综合考虑发送节点与邻居节点之间链路的质量以及邻居节点队列中数据包数量等影响路径质量的因素，将路由选择过程建模为 Markov 决策过程，并使用分布式值函数算法求解当前网络状况下的最优路径。若发送的是低优先级数据，则在路由表中随机选择一条最短路径发送。仿真结果表明，该协议能够有效降低高优先级数据包的端—端延时、提高数据包投递成功率，且对不同流量强度的网络均有一定的适用性。

[参　考　文　献]

[1] 王文光，刘士兴，谢武军. 无线传感器网络概述[J]. 合肥工业大学学报(自然科学版)，2010，33(9)：1416-1419，1437.

[2] MITCHELL R，CHEN I R. A survey of intrusion detection in wireless sensor network applications[J]. Computer Communications，2014，42：1-23.

[3] RADI M，DEZFOULI B，BAKAR K. Multipath routing in wireless sensor networks：survey and research challenges [J]. Sensors，2012，12(1)：650-685.

[4] KOGA Y，SUGIMOTO C，KOHNO R. Congestion control routing protocol using priority control for ad-hoc networks in an emergency[C]//12th International Conference on ITS Telecommunications，Taipei，Taiwan，2012：45-49.

[5] ZHANG Y J，YAN T，TIAN J. TOHIP：a topology-hiding multipath routing protocol in mobile ad hoc networks[J]. Ad Hoc Networks，2014，21：109-122.

[6] 公维冰，阳小龙，张敏. 基于细胞适应机制的自组网多路径路由协议[J]. 通信学报，2014，35(6)：56-63.

[7] CHANAK P，BANERJEE I. Energy efficient fault-tolerant multipath routing scheme for wireless sensor networks[J]. The Journal of China Universities of Posts and Telecommunications，2013，20(6)：42-48.

[8] SADEGHI P，KENNEDY R，RAPAJIC P. Finite-state Markov modeling of fading channels：a survey of principles and applications [J]. IEEE Signal Processing Magazine，2008，25(5)：57-80.

[9] 唐昊，万海峰，韩江洪. 基于多 Agent 强化学习的多站点 CSPS 系统的协作 Look-ahead 控制[J]. 自动化学报，2010，36(2)：289-296.

[10] SHIRAZI G N，KONG P Y，THAM C K. Distributed reinforcement learning frameworks for cooperative retransmission in wireless networks[J]. IEEE Transactions on Vehicular Technology，2010，59(8)：4157-4162.

（责任编辑　张淑艳）