

有效网络大数据流多任务传输调度方法

庞 慧¹, 刘丽娟², 周丽莉¹

(1. 河北建筑工程学院信息工程学院, 河北 张家口 075000;

2. 河北农业大学信息科学与技术学院, 河北 保定 071000)

摘要: 针对现有方法在调度中存在的丢包和数据拥塞的问题, 在有效网络的基础上构建一种大数据流多任务传输调度方法。将无线传感器网络看作有效网络, 构建对应的数据传输网络低占空比模型, 根据模型运行原理分析大数据流传输调度问题。设计网络服务曲线调度框架, 实现传输节点能量最优, 计算由数据拥塞造成的传输延迟函数, 采用能量延迟权衡原则优化数据传输能量效率。结合数据传输调度框架以及能量最优调度问题, 提出 LURS 传输调度算法, 充分考虑多任务传输调度的周期问题, 在每个周期中执行一次能量最优的时间槽分配, 从而实现多任务传输的协调调度。实验结果表明, 所提方法能够降低数据拥塞率, 且任务调度中的数据丢包现象得到了有效缓解。

关键词: 数据传输; 多任务; 低占空比模型; 调度算法; 数据拥塞; 无线传感器网络

中图分类号: TP316 **文献标识码:** B

Effective Scheduling Method for Multi Task Transmission of Big Data Stream in Network

PANG Hui¹, LIU Li-juan², ZHOU Li-li¹

(1. College of Information Engineering, Hebei University of Architecture, Zhangjiakou Hebei 075000, China;

2. College of Information Science and Technology, Hebei Agricultural University, Baoding Hebei 071000, China)

ABSTRACT: Due to the data loss and data congestion in the current method, this paper constructs a multi-task transmission scheduling method for large data streams based on effective network. The wireless sensor network is regarded as an effective network to construct the corresponding low duty cycle model of data transmission network, and the transmission scheduling problem of large data stream was analyzed according to the running principle of the model. The scheduling framework of network service curve was designed to optimize the energy of transmission nodes. The transmission delay function caused by data congestion was calculated, and the energy efficiency of data transmission was optimized via the energy delay tradeoff principle. The LURS transmission scheduling algorithm was proposed by the data transmission scheduling framework associated with energy optimal scheduling problem. The periodic problem of multi task transmission scheduling was fully considered, and the optimal time slot allocation was implemented once in each cycle, thus achieving the coordinated scheduling of multi-task transmission. The results show that the method reduces the data congestion rate and packet loss rate.

KEYWORDS: Data transmission; Multi-task; Low duty cycle model; Scheduling algorithm; Data congestion; Wireless sensor networks

1 引言

通过网络中的数据传输能够实现网络与现实的融合, 而无线传感器网络作为物联网的终端系统, 其可以主动与现实

世界融合。无线传感器网络的传输节点能量有限, 因此数据传输是无线传感器能耗的主要来源, 此外物联网对多任务调度的实时性具有较为迫切的需求, 数据传输方式是影响任务调度效率的因素之一, 目前, 数据多任务传输调度已成为该研究领域中的重点研究课题^[1]。

现阶段, 常用的数据调度方法主要有 Round-Robin 调度方法、最小发送队列数据调度方法和基于双端队列的数据调

基金项目: 河北省创新能力提升计划项目(19450309D)

收稿日期: 2020-04-26 修回日期: 2020-06-02

度方法,上述方法在实际数据传输过程中,没有考虑到多个传输任务有可能会使网络出现负载不均衡的情况,使传输数据在时间与空间上没有得到有效的利用,导致在调度中存在数据丢包和数据拥塞的问题。为了实现全面且有效的资源利用目的,需要针对数据做出相应的规划处理,并根据网络整体运行状况进行优化调度,这样就可以从根本上对网络运行状态进行调度,从而达到数据均衡的效果^[2,3]。

基于多任务数据传输存在的数据丢失和数据拥塞等问题,提出一种有效网络大数据流多任务传输调度方法。该方法首先构建了低占比模型,并且根据其模型描述网络传输数据所存在的问题;然后根据数据传输的问题构建了最优能量的网络服务曲线,再利用能量延迟权衡原则改进其问题,最后提出一种 LURS 调度算法,运用该算法实现多任务的有效调度。仿真结果表明,所提方法在数据保存和数据拥塞率方面具有明显的优势。

2 网络模型构建与问题描述

2.1 数据传输网络低占比模型构建

将无线传感器网络看做有效网络,而无线传感器网络则看作是一个完整的无向图 $G(V, E)$, 该图中 V 和 E 分别描述传感器网络中节点集合与节点之间的链路集合。为了降低网络中节点运行能量,需要构建节点工作的低占比模型,并在此基础上,把节点 V 的运行周期等分成多个运行时段,这样即可以达到在不同工作周期内,节点 V 只需要在同一时段中接收各种数据的目的。 H_v 表示节点 V 的活跃时间。

此外除了接收数据的时段,其时段中的传输节点会处于一种完全睡眠的状态,而这个状态会一直持续到所需发送的数据发完为止。如果这时传感器节点有了对应的处理动作,并且还有相同的节点工作周期 T ,那么每个节点就可以提前知道相邻节点之间的活跃时间。

将节点运行时隙设置为 1,并且将该时段看作最短时段。在实际数据多任务传输的过程中,其中任意的一个传输任务需要针对最原始的指令请求进行反复确认,然后在此基础上根据传输路径将数据发送到该任务的最终目的地,由于网络中传输的大数据流是多任务的,则每个任务就可以描述为 $TASK_i (1 \leq i \leq n)$, $v_{ri}, v_{di}, PATH_i, NODE_i, D_i$, 其中 v_{ri} 和 v_{di} 分别表示最初节点和最终节点, $PATH_i$ 和 $NODE_i$ 表示从 v_{ri} 到 v_{di} 数据传输路径的边和节点, D_i 表示数据传输过程中节点传输的时间控制因素。

节点传输数据的时间调度表 S 记载了传感器节点的数据接收时间,其中将 $S(i, j)$ 描述为节点 $v_j \in NODE_i \setminus \{v_{ri}\}$ 接收的多任务 $TASK_i$, $S(i, j, t_i)$ 表示多任务 $TASK_i$ 的时延^[4]。根据上述分析可知,任务 $TASK_i$ 的时间安排表还可以表示为 $v_{ri} \rightarrow v_{di}(t_{ki}) \rightarrow \dots \rightarrow v_{di}(t_{di})$, 将 t_j 描述为节点 v_j 从数据传输路径上的任意一节点处接收数据的时间,便有 $t_{di} = S(i, d_i)$ 。

图 1 为低占比模型示意图,如图所示,如果在 $v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow v_3$ 路径上的节点,在合理范围内没有进行合法的升降

序时间序列,那么这些数据就不会在规定时间内将数据发送到 v_3 中。

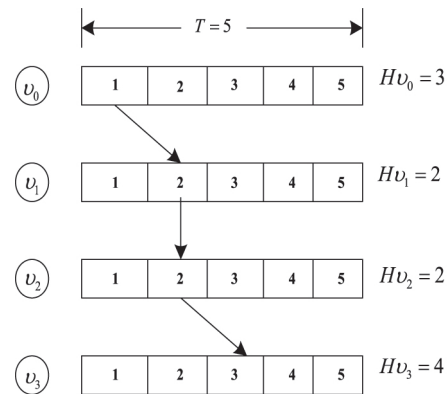


图 1 低占比模型示意图

2.2 传输问题描述

在网络正常运行的情况下,由于多任务数据同时进行传输,会导致网络出现负载不均衡的情况,所以数据传输调度就是指在多个任务同时执行的过程中,让网络运行可以达到均衡的标准,将传输过程中存在的问题描述为:令网络中存在的节点传输周期设定为 T ,并且将数据传输过程中的单跳时间^[5]控制为 p ,那么节点活跃时间就为 $H_v, \forall v \in V$ 。此时如果网络节点接收到 n 个传输数据流的任务指令,做出假设 $1 \leq i \leq n$,那么便有 $D^* = \max \{D_i\}$,并且 $w(j, k) = \sum_{i=1}^n x(i, j, k)$,其中节点传输数据目标是实现网络传输吞吐量的最大化

$$W(S) = \max \left(\sum_{k=1}^{D^*} \sum_{i=k}^{i+p} x(i, j, k) \right) \forall v_i \in V \quad (1)$$

将 V 描述为节点 v_j 在固定时间内的网络整体负载情况。此时,网络中所有传输节点中的部分节点只有一次机会接收数据,其中最大负载为 2 的调度表就可以写为

$$\begin{aligned} v_1 \rightarrow v_3(3) \rightarrow v_5(3) \quad v_2 \rightarrow v_3(8) \rightarrow v_5(8) \\ v_0 \rightarrow v_3(3) \rightarrow v_6(5) \quad v_4(8) \rightarrow v_6(5) \end{aligned} \quad (2)$$

通过上式即可得到 $v_2 \rightarrow v_3(3) \rightarrow v_5(3)$,那么网络整体负载情况将会上升至 3。

3 多任务传输调度

3.1 最优传输调度框架

根据上述问题描述,将数据多任务调度问题合理转变为传输节点能量最优网络服务曲线^[6]构建的问题。在该系统中,传感器节点会周期性地形成各种不同的调度数据包,并且会在该系统将调度数据包进行存储。

基于上述问题,将无线传感器网络传输数据的服务请求函数判定为 $A(t) = \sum_{i=1}^N A_i(t) (t \in [0, T + d_0])$,其中 $A(t)$ 表示一个单调递增的右连续阶梯函数,传输延迟函数^[7]就

可以被判定为 $C(t) = \sum_{i=1}^N C_i(t)_i$ 。假设网络服务曲线为 $(SC) S'(t)$ 其函数需要满足以下三个条件:

- 1) $S'(0) = 0$;
- 2) $S'(T+d_0) = A(T^+)$;
- 3) $C(t) \leq S'(t) \leq A(t) \quad (0 \leq t \leq T+d_0)$ 。

针对节点多任务数据传输的调度问题,将累计流量函数设定为 $A_i(t) \quad (i=1, 2, \dots, N)$, 则调度数据包 B_{ij} 就可以通过计算获取出实际取值,在此基础上可以获取下述关系定理。

定理: 针对任意一个同时满足于各种约束条件的时分传输调度 $\{T_{ij} | i=1, 2, \dots, N; j=1, 2, \dots, M\}$, 至少存在一个网络服务曲线 $S'(t)$ 与该网络服务曲线 $S(t)$ 对应的时分传输调度可以表示为

$$\{T_{ij} | i=1, 2, \dots, N; j=1, 2, \dots, M\} \quad (3)$$

证明: 结合上述定理进行判定, 其中网络服务曲线 $S'(t)$ 按照以下方式进行构造。当 $1 < i \leq N$ 时, 那么对于任意一个条件函数 $j=1, \dots, M$, 都可以用直线段连接点 (t_{ij}, b_{ij}) 和 $(t_{i-1}, b_{i-1,j})$ 进行表示, 具体内容如图 2 所示。

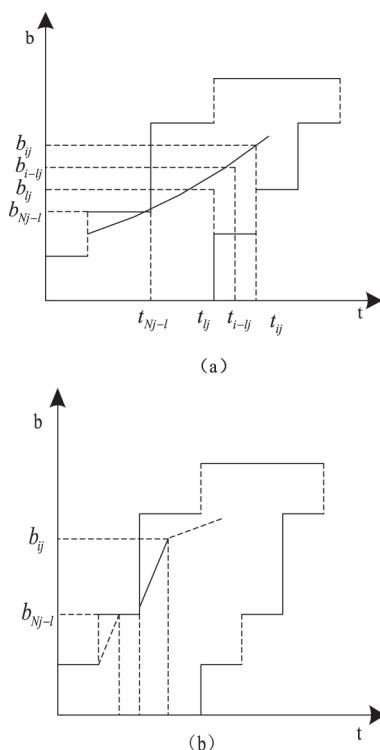


图2 网络服务曲线示意图

根据上图所示, 当 $i=1$ 时, 对于任意 $j=2, \dots, M$, 如果 $t_{Nj-1} > (j-1)\Delta d$ 那么可以用直线段连接 (t_{1j}, b_{1j}) 和 $((j-1)\Delta d, b_{Nj-1})$ 两点, 如上图 2(a) 所示。否则连接 $((j-1)\Delta d, b_{Nj-1})$ 和 (t_{1j}, b_{1j}) 两点, 然后再用直线段连接点 $((j-1)\Delta d, b_{Nj-1})$ 和 (t_{1j}, b_{1j}) , 如上图 2(b) 所示。

根据上述分析可知, 由于网络中节点在接收数据的过程中, 并不会因为时间的增多而减少数据接收量, 所以网络服

务曲线 $S'(t)$ 在计算中是一种单调递增的连续函数, 假设 $S'(t)$ 在逐渐递增的过程中没有从单调增加方面执行, 即可得出时间区域 $(t_1, t_2) \subset [0, t]$, 并且对于 $\forall t_3 \in (t_1, t_2)$, 有 $S'(t_1^+) = S'(t_3) = S'(t_2^-)$, 但是在这种情况下会间接导致网络中簇头节点在对数据传输任务进行分配时, 在没有接收任何数据的情况下, 会出现在其时间区间中, 针对此种情况采用能量延迟权衡原则^[8], 进行综合计算即可让网络服务曲线的能量效率得到有效的提升。

3.2 大数据流下多任务传输调度算法

结合上述数据传输调度框架, 针对网络大数据传输能量的优化调度问题, 提议出一种 LURS 传输调度算法, 首先考虑如何在线划分广播时间, 在广播时间内, 每个传感器节点将数据包传输到簇头节点, 由于网络服务需求曲线和传输延迟曲线都是未知的, 由于没有新的网络服务请求进入队列, 因此, 在此时间间隔内, 能量转移规划的最优策略保持不变。基于此提出了在线传输计划算法^[9], 以确保在采样间隔内传输终端是能量最优的, 该算法在实际计算过程中, 可以针对网络能量效应进行重新计算, 并且根据结果即可得知目前网络多任务传输数据的最优方法。

其中数据传输能量-速率函数即可写为

$$P_i(r) = e_i P_0(r) \times T_{ij} \quad (i=1, \dots, N) \quad (4)$$

式中 $P_0(r)$ 表示网络服务速率 r 的凸递增函数, e_i 表示传感器节点 v_i 的能量扩展系数。

此处做出假设, 将 l_i 描述为传感器节点 v_i 和簇头节点之间的固定距离, 则有 $e_i \propto N_i l_i$ 。由于所有传感器节点在一个调度周期内均被调度过一次, 那么可以得到平均能量的速率函数^[10]:

$$\bar{p}(r) = \frac{1}{N} \sum_{i=1}^N e_i P_0(r) = \bar{e} P_0(r) \quad (5)$$

式中 $\bar{e} = \frac{1}{N} \sum_{i=1}^N e_i$ 。

根据上式计算结果即可得知 $\bar{p}(r)$ 也是一个递增函数, 应用恒速能量优化原理, 使局部能量在一个规划周期内得到最优的规划策略。

根据上式计算即可进一步使数据传输节点在有效网络中所形成的数据包得到一定的缓存处理, 并在此基础上根据求解结果获取出当前节点能量优化的调度算法。随后在确定了能量优化的服务曲线之后, 即可计算得出每个调度数据包被分配到的时间窗口, 那么与其对应的时隙即可在计算过程中通过调度周期开始执行调度。

由于现有方法在实际计算的过程中只实现了传输调度工作周期分化的目的, 没有在能量最优的前提下实现调度周期转换为时间槽的目的, 针对该问题, 根据在每个周期中执行一次能量最优的时间槽分配, 对 LURS 传输调度算法进行改进。

在实际计算中, 根据下式即可获取出与其相对应的传输调度时间槽:

$$\min \sum_{i=l_0}^N E_i = \sum_{i=l_0}^N \frac{T_{ik_0}}{|h_i|^2} (2^{B_{ik_0}} - \bar{p}(r)) \quad (6)$$

$$\sum_{i=l_0}^N T_{ik_0} = S^{-1} \left(\sum_{i=1}^N \sum_{j=1}^{k_0} B_{ij} \right) - k\Delta d \quad (7)$$

式中 $T_{ik_0} > 0 \quad i = l_0, \dots, N$ 。

4 仿真研究

为了能够验证所提有效网络大数据流多任务传输调度方法的实际有效性,将在仿真中采用 TOSSIM 模拟器,在 $100\text{m} \times 100\text{m}$ 方形区域上随机布置 30 ~ 100 个传感器节点,使用默认发射功率,采用 Matlab 软件进行数据处理,并将所提方法与 Round-Robin 调度方法、最小发送队列数据调度方法和基于双端队列的数据调度方法进行对比。

当对每个传输任务的分组数量进行改变时,与其相对应的数据拥塞率会随之产生变化,当数据拥塞率有所提高时,有效网络便会出现负载不均衡的情况,数据分组丢失和缓冲溢出便会呈现上升的状态。在此状态下进行多任务调度,不同方法在多任务调度中的数据拥塞率对比结果如图 3 所示。

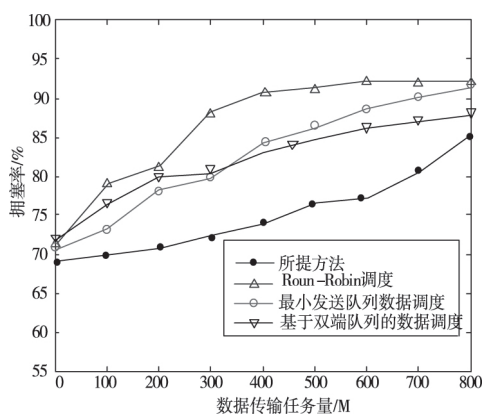
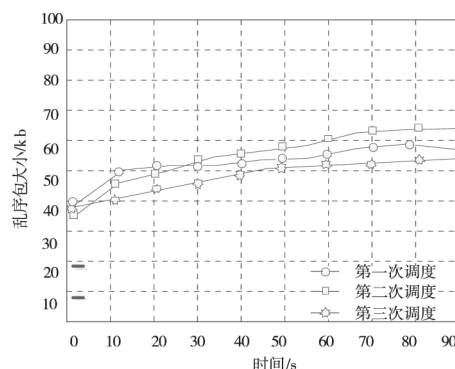


图 3

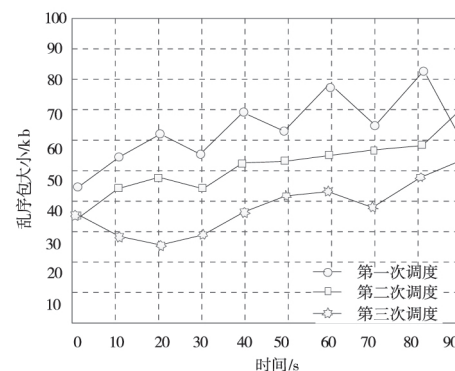
分析图 3 可知,本文所提出的调度方法造成的拥塞率明显低于现有方法,其最低值为 68%,说明运用该方法在有效网络中进行大数据流多任务传输调度时,数据传输更加流畅,能够达到理想水平。这是因为所提方法通过计算由数据拥塞造成的传输延迟函数,采用能量延迟权衡原则优化了数据传输能量的效率,从而降低了数据拥塞率。

在多任务传输中,会产生大量的乱序包,乱序包会造成网络拥塞,影响用户体验效果,为此将乱序包作为检验多任务传输调度方法的指标,对不同方法进行对比,结果如图 4 所示。

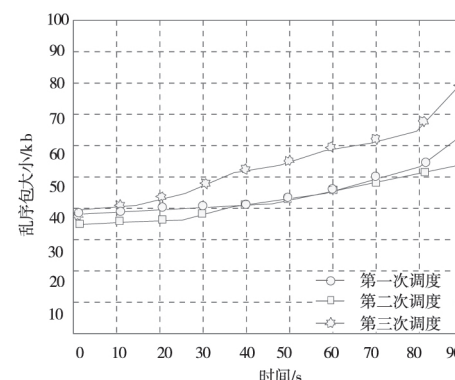
分析图 4 可知,在三次传输调度中,不同方法下的乱序包大小不同,其中,所提方法的乱序包大小明显低于 Round-Robin 调度方法、最小发送队列数据调度方法和基于双端队列的数据调度方法,说明所提方法能够有效降低数据丢失量,与传统方法相比具有十分显著的优势性。这是因为所提



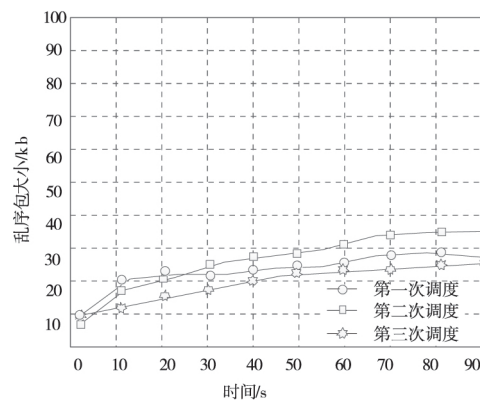
(a) Round-Robin调度方法



(b) 最小发送队列数据调度方法



(c) 基于双端队列的数据调度方法



(d) 所提方法

图 4 不同方法的乱序包大小对比

方法充分考虑了多任务传输调度的周期问题,分别在每个周期中执行一次能量最优的时间槽分配,从而减低了数据丢失的概率。

5 结论

现阶段随着网络数据应用的快速发展,有效网络中传输数据的负载能力也随之成为了该领域中重点研究的课题之一,但由于传统方法在调度过程中存在数据丢包和网络拥塞率较大的问题,基于此本文提出有效网络大数据流多任务传输调度方法。首先构建网络数据传输调度模型,并且根据网络模型针对节点传输数据调度问题进行描述,随后利用能量延迟权衡原则针对数据传输能量效率进行改进,并通过LURS调度算法实现有效网络中大数据流多任务传输调度的目的。经实验结果证明,所提方法可以从根本上解决传统方法带来的问题,并且具有实用性强的优点。

参考文献:

- [1] 郝春亮,沈捷,张珩,等. 大数据背景下集群调度结构与研究进展[J]. 计算机研究与发展, 2018, 55(1): 53-70.
- [2] 张本宏,王一茗,俞磊,等. 基于数据到达率的IWSNs分层调度方法[J]. 电子测量与仪器学报, 2019, 33(6): 76-81.
- [3] 薛娟,高红伟,姜辉,等. 基于时隙ALOHA协议的数据传输二人随机博弈模型[J]. 运筹学学报, 2019, 23(4): 45-58.

(上接第365页)

- [3] 包伟涛,韦钢,刘佳,等. 基于区间模型的多能互补系统优化运行[J]. 电力系统自动化, 2019, 43(11): 8-16.
- [4] 朱晔,兰贞波,魏震,等. 考虑碳排放成本的风光储多能互补系统优化运行研究[J]. 电力系统保护与控制, 2019, 47(10): 127-133.
- [5] 肖林,刘丽芳,卫星,等. 钢-混组合索塔锚固结构的力学行为及结构优化[J]. 西南交通大学学报, 2019, 54(5): 923-930.
- [6] 张锡治,李青正,章少华,等. 钢-混凝土预制混合梁变形性能研究[J]. 工程力学, 2019, 36(6): 193-201.
- [7] 刘丽萍,李坤,田慧欣. 钢铁企业出厂物流集成多目标优化问题[J]. 控制工程, 2019, 26(3): 502-509.
- [8] 瞿萧羽,陶进. 寒地公共建筑取暖能耗总量准确预测仿真[J]. 计算机仿真, 2018, 35(5): 379-382, 425.
- [9] 马国栋,汪梦甫. 耐震时程法在钢-混混合结构中的应用[J]. 公路工程, 2019, 44(4): 48-53.

- [4] 彭鑫,邓清勇,田淑娟,等. 多信道车联网V2R/V2V数据传输调度算法[J]. 通信学报, 2019, 40(3): 92-101.
- [5] 肖宇,徐任晖,朱军,等. 一种定向传输武器协同数据链的时分调度信息共享方法[J]. 通信技术, 2019, 52(6): 1391-1396.
- [6] 于琨,张正本,海本斋. 基于多目标全局约束的任务分配和调度算法[J]. 计算机工程与应用, 2018, 54(8): 55-60.
- [7] 徐飞,王少昌,杨卫霞. 基于博弈论的云资源调度算法[J]. 计算机科学, 2019, 46(1): 295-299.
- [8] 杜鹏,晏亮,高保成,等. 基于电力调度数据网的广域数据采集方案[J]. 电力系统自动化, 2019, 43(13): 156-162.
- [9] 陈建彪. 基于机器学习的复杂网络数据流均衡调度仿真[J]. 计算机仿真, 2019, 36(11): 264-267.
- [10] 刘义才,刘斌. 具有时延和丢包的网络化切换系统控制研究[J]. 控制工程, 2018, 25(08): 1482-1489.



【作者简介】

庞 慧(1979-),女(汉族),河北张家口人,硕士,副教授,主要从事计算机系统、大数据、物联网方向的研究。

刘丽娟(1980-),女(汉族),河北保定人,硕士,副教授,主要从事数据分析、数据存储方向的研究。

周丽莉(1967-),女(汉族),河北张家口人,硕士,教授,主要从事数据库、数据挖掘、大数据方向的研究。

- [10] 张沁宜,林波荣. 典型绿色办公建筑能耗差异校验与敏感性分析[J]. 暖通空调, 2019, 49(8): 31-39.
- [11] 吴会来,谭洪卫,邓丰. 基于实测数据的夏热冬冷地区低密度住宅近零能耗技术实践研究[J]. 建筑科学, 2019, 35(6): 1-8, 17.
- [12] 田东,张士兴,幸国权,等. 装配式混凝土建筑与超低能耗技术应用研究[J]. 建筑技术, 2019, 50(8): 918-920.



【作者简介】

李丹丹(1981-),女(汉族),吉林公主岭人,硕士研究生,副教授,研究方向:建筑节能。

李 健(1984-),男(汉族),辽宁大石桥人,硕士研究生,工程师,研究方向:建筑节能。

陶 进(1958-),男(汉族),吉林长春人,硕士研究生,教授,研究方向:建筑节能。