

A Large-Scale Data Collection and Processing Framework for Wireless Sensor Networks*

FAN Lei, GONG Wentian, SHI Xiaoqiu*

(College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou Zhejiang 325035, China)

Abstract: Data collection and processing is an important component of wireless sensor network applications. In this work, we introduce a new system design framework that includes large-scale sensor data collection, information extraction, storage and online publishing. In our framework, the sensor data gathering to sink node are first real-time batch processed by using an independent thread; and then the data are stored in relational databases specially designed; finally, the network topology and multivariate spatio-temporal data visualization are provided based on B/S mode for interactive remote monitoring. Experimental result in large-scale and long-term WSN scenario shows that this framework meets the data collection and processing requirements about real-time, reliability and usability, and its universal flexible design has advantages of expansibility and portability, which would be of benefit to more WSN applications.

Key words: wireless sensor networks; data collection and processing; relational schema design; B/S; visualization

EEACC: 6150P

doi: 10.3969/j.issn.1004-1699.2011.11.020

大规模无线传感器网络数据收集与处理系统设计*

樊 镭, 龚闻天, 施晓秋*

(温州大学物理与电子信息工程学院, 浙江 温州 325035)

摘 要: 数据收集与处理是无线传感器网络应用系统的重要组成部分。本文给出了一种集传感器数据实时收集、信息提取与存储、以及在线发布功能的系统设计方案, 汇聚至 Sink 结点的传感数据由上位机收数模块采用独立线程进行批量收取、拆分、错误检测、翻译等处理后存入关系数据库, 基于 B/S 模式提供了网络拓扑结构和多元时空数据的交互式可视化远程监测。实验表明, 系统在实时性、可靠性和可用性方面满足大规模长期部署的无线传感器网络应用需求, 其通用灵活的设计架构具有较高的可移植性和可扩展性。

关键词: 无线传感器网络; 数据收集与处理; 关系模式设计; 浏览器/服务器模式; 可视化

中图分类号: TP393

文献标识码: A

文章编号: 1004-1699(2011)11-1611-06

1 问题与挑战

无线传感器网络 (Wireless Sensor Networks, WSNs) 作为 Ad-hoc 网络的一种典型应用, 近年来已经广泛部署在如火山^[1]、海洋^[2]、野生动物栖息地^[3]、森林环境监测^[4]、煤矿救援^[5]以及精准农业^[6]等多种领域。与传统互联网不同, 无线传感网络可以看作一种获取数据的方法, 传感器结点本身就是数据源, 周期性的对环境参数进行采样并通过多跳路由把数据传回到一个或多个汇聚结点 (Sink)。无线传感器网络项目的成功实施应该包括

两个部分: ①部署结点; ②收集数据。通信和网络方面的研究人员通常更关注前者, 即组网传输数据, 目前在 MAC 层通信、拓扑控制、时间同步等相关技术方面已经做出了很多卓越的工作。虽然部署与组网是无线传感器网络的基础, 但是在实际应用中, 仅仅一个独立封闭的 WSN 通信网络是没有意义的, 最终目的仍然是能够把网络中监测的数据收集起来, 经过上层服务处理后提供给终端用户使用。因此, 有效的数据收集与处理作为底层网络与上层应用之间承上启下的桥梁, 是整个无线传感器网络应用中非常重要的组成部分。

然而,由于 WSNs 具有自组织、多跳传输、监测数据多元时空关联、一经部署很难更改等特点,使得有效的数据收集和处理面临较大挑战。对于大规模长期部署的无线传感器网络应用场景,首先, Sink 结点在收取大量汇聚数据时存在性能瓶颈,如文献[7]中通过在传感器结点中使用嵌入式 Web 服务器进行数据的采集和发布,用户能够直接通过浏览器监测传感器数据,但对于资源有限的传感器结点来说这种方式开销大,且效率较低无法满足性能需求,不适合大规模结点部署和大数据量处理;其次, WSN 长期增量部署过程中,面临如何灵活高效的适应有新的异构结点加入时所产生的数据存储和查询等问题,文献[8]给出的方案能够实时处理测试组网中 10 个结点的数据,但其将收数程序、服务程序以及数据库存储的实现全部集中在嵌入式网关,功能扩充与软件升级缺乏足够的灵活性,可扩展性较差,网关易成为网络的单故障点,且基于 C/S 的监测模式需要安装专门的客户端软件,不利于数据发布和共享。再次,在需要长期大规模部署的情况下,动态的网络拓扑和廉价的传感器结点等不稳定因素使得系统可靠性和可扩展性难以保证,而采用复杂的网内数据融合、嵌入式数据处理等技术进一步增加了网内的计算、通信、存储、能量等开销,缩短了整个无线传感器网络的生存期^[1,9];另外无线传感器网络收集的数据最终必须能够被方便的查询和使用才有意义,而现有文献都没有详细涉及对 WSN 网络拓扑结构和多元时空数据流的可视化处理方法,而这对于提高系统远程监测的可用性也是非常必要的。

针对上述问题,本文给出了一种适合大规模长期部署的无线传感器网络应用的数据收集与处理解决方案,并将基于该方案开发实现的系统实例 Receiver 用于基于 WSNs 的杨梅生长环境生态监测任务,取得了良好的效果。下文第 2 节详述了系统架构与各模块设计,第 3 节给出了系统具体实现,第 4 节介绍了系统运行在真实环境中的实验测试结果,第 5 节对全文进行了总结。

2 系统设计

数据收集处理系统包括收数、存储与发布三大模块,图 1 给出了与上述功能所对应的系统总体架构。该架构采用了集中式设计,简洁高效,接口清晰,实现了对传感器数据的快速准确收取、灵活持久存储以及多元时空数据的可视化等功能;尽可能降低与底层网络硬件、通信协议的耦合性,以便较好的

适应今后网络中异构结点的加入;各模块之间采用 TCP/IP 通讯协议传递数据,部署灵活,且容易扩展和移植。汇聚到各个 Sink 结点的网络包首先由收数模块采用独立线程实时监听并批量收取,然后经过抽取、错误分析、翻译等必要处理,根据包的信息结构拆分存储到关系数据库相应表中,最后基于 B/S 模式和 AJAX 技术的 Web 在线发布模块对存储模块中的网络拓扑结构和多元时空数据等信息进行可视化处理,供终端用户使用。

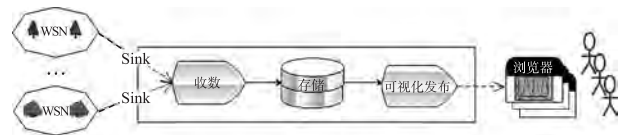


图 1 系统架构

采用集中式的设计方案主要基于以下考虑:

(1)性能开销:数据收集处理往往需要大量的运算资源,而廉价传感器结点资源非常有限,分布式的数据处理设计必然增加结点的计算、存储、通信等开销,并增加结点能耗^[9],导致系统生存期缩短。集中式处理可以避开这种限制,通过利用上位机端资源,有效减少无线传感器网络内部开销,同时提高了收集与处理的效率。

(2)灵活性:与嵌入在结点中的分布式处理方法相比,集中式设计不对无线传感器网络的底层部署做任何假设和限制,降低了各功能模块之间的耦合度,可以有效的与底层实现分离,更好的适应不同组网机制。而分布式设计通常依赖于具体的硬件、嵌入式操作系统、MAC 层协议、拓扑控制等实现,不利于扩展和移植。

(3)可靠性:无线传感器网络系统的部署经验表明,在网络内部引入复杂的处理机制会导致系统行为难以理解、程序难以调试,除错及修复的复杂度增加,应该使网络内部的处理尽可能简单,以最大程度的确保无线传感器网络系统成功实施。将数据集中在上位机端进行处理,仅把网络中的传感器结点当作采集数据的从属设备,简化了网络内部的设计,有效克服了分布式处理中易造成网络不稳定的弊端,使得系统功能扩充和升级更加容易,即使个别传感器结点的故障和失效也不会影响无线传感器网络其他部分的正常运行,增强了系统在大规模长期部署应用中的可靠性和可用性。

下面分别介绍系统各组成模块中关键技术的详细设计。

2.1 数据批量实时收集处理

收数模块位于系统的第一层,由监听、批量收

取、抽取、错误分析检测和翻译等功能组成,性能上确保快速准确。图 2 给出了收数模块的流程设计。

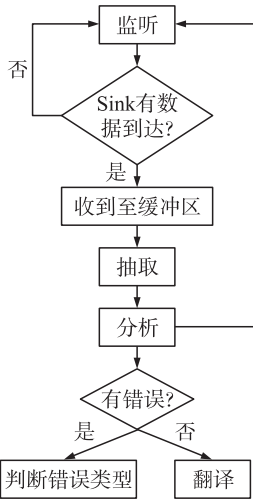


图 2 收数模块处理流程

程序启动初始化后,创建独立的线程持续监听 Sink 结点的汇聚状态,当有数据到达时,触发收取事件将本次数据放入缓冲区。由于采用批量方式,一次收取的数据可以由多个包组成,需要根据包头标识抽取单独的原始包;数据收取后,通过校验原始包的数据长度、固定位置的值等预定义规则来分析原始包是否出现错误,若有错误则记录错误类型代码,否则由翻译函数按照每个原始包的结构格式判断所属类型,转换得出包中各个字节代表的属性值,并将这些处理后得到的结果交给存储模块按相应设计予以保存。批量收取并拆分的方式能够明显改进处理大量数据时的效率,同时该过程中得到的错误类型信息可被后面的数据发布模块用来决定是否排除,或者供研发人员作为排查错误原因的线索,以便及时发现潜在的网络问题。

2.2 基于结构化信息的存储设计

存储模块为数据收集处理系统的中间层,负责永久保存上一阶段收集处理后得到的数据结果,为上层功能模块提供数据源服务。文件形式存储是最原始的方法,即按时间段将数据直接写入 TXT、CVS、XML 或 Excel 等格式的文件,虽然实现简单但不利于并发查询与处理。将无线传感器网络数据存储于关系数据库能够充分发挥存储模块的作用,但需要合理的设计才能保证灵活高效。基于收数模块分析翻译处理后得到的原始包结构化信息,图 3 给出了相应的概念模式设计。

根据无线传感器网络中原始数据包的信息组成结构特点拆分建立相应的存储表,所有表都建立一个自增字段作为逻辑主键,在被参照时作为外键使

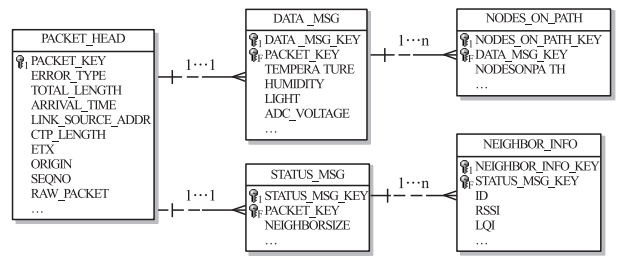


图 3 存储模块关系数据库概念设计

用,保证关系模式的实体完整性约束和参照完整性约束。其中包头信息表、监测数据信息表、状态信息表、路由结点表和邻居结点表之间按上下文联系建立一对一或一对多的参照关系。整个关系模式设计符合 BC 范式要求,消除冗余且保持查询效率和扩展的灵活性,这在今后应对新类型的数据加入时非常便利,只要为其特有的信息创建新表即可,其他共同的信息部分可以重用现有设计,使得数据收集模块程序和数据库关系模式的修改都大大减少;即使今后部署使用不同嵌入式操作系统和路由协议的异构网络结点,只要收集的数据信息类似,也可以按照属性值存入相应的表和字段,达到数据整合的目的。

2.3 网络状态和多元时空监测数据的可视化处理

无线传感器网络收集到的数据流是监测数据在时间和空间上的分布。数据发布模块针对这种特点,给出了一种可视化的查询分析方案,着重对随时间变化的 WSN 网络拓扑结构等状态信息以及多元时空关联的传感数据信息进行处理,提高了系统的可用性。具体功能设计结构如图 4 所示(图中实线和虚线分别表示数据支持和可视化支持,方框表示空间相关,圆框表示时间相关,圆角方框表示时空相关)。

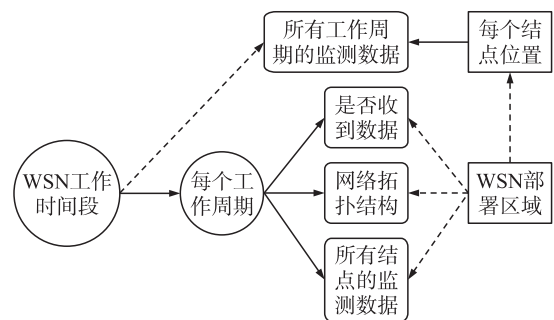


图 4 可视化发布模块功能设计

由于 C/S 模式中专门的桌面监测软件对用户平台有一定限制要求,且客户端安装和维护的工作量大,不利于无线传感器网络数据的监测和使用,因此本系统的数据库发布模块基于 B/S 模式设计。多元时空信息的在线交互式 Web 页面采用 AJAX (Asynchronous JavaScript And XML) 技术进行可视化处理。相对于 Flash 和 Java applet 等交互式技术,AJAX 是一种使用

户操作与服务器响应异步化的轻量级解决方案,不需要用户安装额外的插件或运行环境,且遵循 W3C 标准,有很好的开放性、兼容性和可移植性。另外为了能够在浏览器端及时更新数据,Web 程序中加入定时器后台实时检查数据收集状况,当发现有新数据时自动更新页面显示并给用户发出提醒信息,达到实时监测的效果。数据发布模块将以上网络部署状态信息和传感器监测数据信息集成在一起,任何连接到 Internet 的终端用户都可以简单的通过浏览器远程在线访问,并能够通过可视化环境动态交互的查询和分析收集的数据,使无线传感器网络真正走出实验室,在实际应用中发挥其特有的优势。

3 系统实现——Receiver

基于上文的系统架构设计所实现的系统实例名为 Receiver。其中收数模块根据 Sink 结点传送原始数据包至上位机的方式,如 RS-232 串口或 GPRS 等,使用 Delphi 语言多线程监听串口或网络端口编程实现;存储模块中的表、存储过程、触发器、视图等概念模式使用开源的 Firebird 关系数据库实现,目前共同运行在一台 Windows 操作系统的 PC 机上,专门用于收集和存储数据,分配内网 IP 为上层应用提供服务。图 5 给出了 Receiver 系统收集数据过程中的用户界面截图。

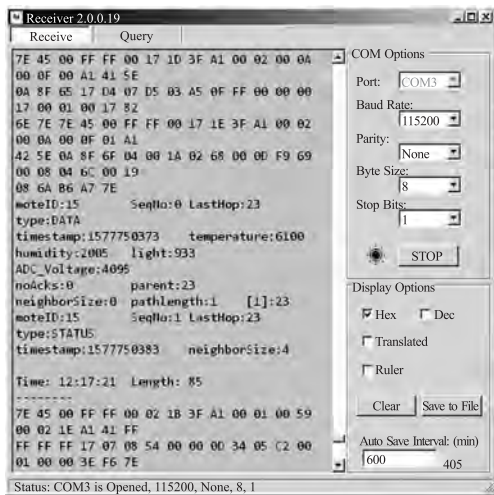


图 5 Receiver 收数模块工作界面

Receiver 系统基于 B/S 的可视化发布模块的服务端程序使用 PHP 语言编程实现,部署在一台配有 PHP 和 Nginx 环境的 Linux 服务器上,分配了独立 IP 和域名,通过 TCP/IP 连接 Firebird 数据库请求数据,对外提供 Web 服务 (http://network.wzu.edu.cn/wsn/demo/)。浏览器端对于网络拓扑结构和多元时空监测数据的 Web 在线交互式可视化处理

分别实现如下:工作周期属于时间数据,通过下拉列表显示;部署区域和结点位置属于空间数据,调用 Google Maps API 中的地图接口实现;每个工作周期的活动结点和网络拓扑结构等网络状态信息属于时间属性约束下的空间数据,分别调用 Google Maps API 中的标记 (Marker) 图层和折线 (Polyline) 图层显示处理;传感器结点的温度、湿度、光照、电压等监测数据,按照时间和空间两种角度分片得到“每个工作周期内所有传感器结点的监测数据”和“每个传感器结点在所有工作周期的监测数据”,分别调用 Google Visualization API 中的运动散点图 (Motion Chart) 和时序说明图 (Annotated Time Line) 实现可视化查询分析等功能。终端用户在浏览器上监测访问的界面截图如图 6 所示。

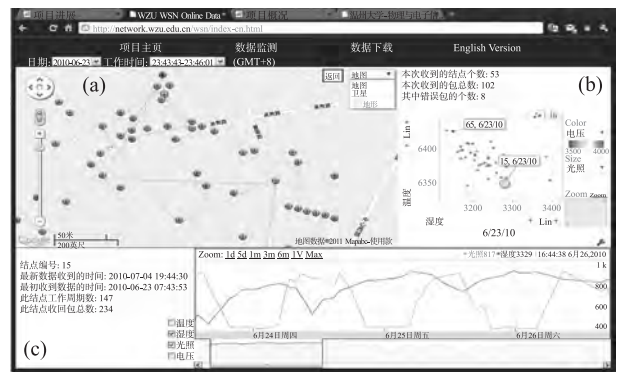


图 6 Receiver 可视化发布模块页面

Receiver 系统收数、存储和发布三个组成部分的部署非常灵活,每个模块分别占用一台计算机或集中部署在同一台计算机都是可以的,实际应用中可以根据情况任意组合,只要模块之间能够通过 TCP/IP 通信即可。

4 实验结果

以基于 WSN 的杨梅园环境监测项目为实验平台,如图 7 所示,在实验基地约 27 000 m² 区域内 (北纬 27°55'5", 东经 120°41'32") 部署了 100 个左



图 7 杨梅生长环境生态监测部署

右传感器结点,通过持续测量温度、湿度、光照、二氧化碳等相关参数实现对杨梅生长环境的长期实时监测。原型系统底层组网使用的是运行 TinyOS 2.1 嵌入式操作系统和 CTP 路由协议^[10]的 TelosB 结点,装配有温度、湿度、光照等传感器,部署封装中使用两只干电池供电。对于环境监测应用而言,采样率过于频繁的传感器测量值一般属于冗余数据^[11],所以为了减少网内的数据量和通信量,改善结点能耗以延长网络生存期,采用 Low-duty-cycle 周期性同步“休眠—工作”的方法^[4,12]对监测数据进行采样和传输。而这种工作方式也使得每次会有大量网络包集中在很短时间内突发性的汇聚到 Sink 结点。

收数模块的批量收取超时的间隔设为 100 ms,即只要两个字节到达的时间间隔小于 100 ms 都被认为是同一帧数据,否则作为下一帧收取。长期测试结果表明每一帧数据从触发批量收取开始,经过拆分抽取、错误检测、翻译等处理后批量存入数据库的延迟一般不超过 300 ms,满足对上述大量突发性数据进行实时收集的需求。表 1 显示了 Receiver 系统的在 6 个月 WSN 增量部署中的数据持续收集情况统计,在总计超过 70 万个网络包中,收取的错误率低于 1%。实验结果表明 Receiver 运行稳定高效,能够适用于长期大规模部署的无线传感器网络应用。

表 1 Receiver 系统数据收集统计

部署时间	最多 结点数	Receiver 收集 的原始包总数	错误 包数
2010 年 2 月	32	46 055	302
2010 年 3 月	70	96 397	727
2010 年 4 月	79	113 006	890
2010 年 5 月	102	136 880	1 027
2010 年 6 月	131	173 725	1 630
2010 年 7 月	125	170 051	1 605
...

终端用户在线监测与查询分析在图 6 页面上方的日期和时间下拉框用于选取工作周期,当前显示的是 2010 年 6 月 23 日 22 点 43 分左右的情况。左上方(a)的地图内显示了部署于杨梅园实验基地区域中的部分结点以及当前工作周期的活动结点(高亮显示)和网络路由拓扑结构。右上方(b)显示本次工作周期数据收集的统计信息以及所有传感器结点监测数据的交互式散点图,其中横轴为温度值,纵轴为湿度值,颜色表示电压值,数据点的大小表示光照值(具体显示模式可调整)。可以看到当时 15 号结点的光照值和电压值明显异常,经检查发现 15 号

结点位于部署区域边缘紧邻校园外的公路,夜晚明亮的路灯和对面商业区的广告霓虹灯使得此结点的光照值在夜间也明显高于其他结点;而较低的电压值提醒监测人员应及时更换电池。点击地图中的某一传感器结点图标,页面下方(c)出现此结点所有工作周期的数据收集情况统计以及时间序列上的监测数据曲线图表,可以看到温度和湿度这两条随时间变化的曲线基本上呈现一种反比的关系,这和经验以及当时的气象情况都是相符的。

5 结语

无线传感器网络作为获取数据的一种手段正越来越多的应用于我们生活的世界,随着组网技术的成熟,对大规模无线传感器网络状态信息和监测数据的有效收集与处理尤为关键。本文提出了一种集数据收集、存储、发布以及可视化查询分析的完整设计与实现方案,基于该方案的实例 Receiver 系统在大规模无线传感器网络长期部署测试中逐渐形成和完善,成功应用于实际监测任务并取得了较好的效果。进一步工作如数据分析、异常检测、事件报警等功能的开发和整合可以在该系统方案的基础上继续进行。另外,该系统架构的通用性和灵活性使其能够很容易移植和扩展,从而在更多的无线传感器网络应用中发挥重要作用。

参考文献:

- [1] Allen G W, Lorincz K, Welsh M, et al. Deploying a Wireless Sensor Network on an Active Volcano [J]. IEEE Internet Computing, 2006, 10(2): 18-25.
- [2] Liu K, Li M, Liu Y, et al. Passive Diagnosis for Wireless Sensor Networks [C]//Abdelzaher T F, Martonosi M, Wolisz A, et al. Proceedings of the 6th ACM International Conference on Embedded Networked Sensor Systems (SenSys). Raleigh: ACM, 2008. 113-126.
- [3] Szewczyk R, Mainwaring A, Polastre J, et al. An Analysis of a Large Scale Habitat Monitoring Application [C]//Stankovic J A, Arora A, Govindan R, et al. Proceedings of the 2th ACM International Conference on Embedded Networked Sensor Systems (SenSys). Baltimore: ACM, 2004. 214-226.
- [4] Mo L, He Y, Liu Y, et al. Canopy Closure Estimates with GreenOrbs: Sustainable Sensing in the Forest [C]//Culler D E, Liu J, Welsh M, et al. Proceedings of the 7th ACM International Conference on Embedded Networked Sensor Systems (SenSys). Berkeley: ACM, 2009. 99-112.
- [5] Li M, Liu Y. Underground Coal Mine Monitoring with Wireless Sensor Networks [J]. ACM Transactions on Sensor Networks (TOSN), 2009, 5(2): 1-29.
- [6] Langendoen K, Baggio A, Visser O. Murphy Loves Potatoes: Experiences from a Pilot Sensor Network Deployment in Precision Agriculture [C]//20th International Parallel and Distributed

- Processing Symposium(IPDPS). Rhodes Island;2006. 8-16.
- [7] 何鹏举,戴冠中. 嵌入式 Web 网络传感器的实时数据采集方法[J]. 传感技术学报,2006,19(6):204-215.
- [8] 张杰,石为人,涂巧玲. 基于无线传感器网络的信息采集监测系统[J]. 传感技术学报,2009,22(6):861-864.
- [9] Allen G W, Haggerty S D, Welsh M. Lance: Optimizing High-Resolution Signal Collection in Wireless Sensor Networks[C]// Abdelzaher T F, Martonosi M, Wolisz A, et al. Proceedings of the 6th ACM International Conference on Embedded Networked Sensor Systems(SenSys). Raleigh: ACM,2008. 169-182.
- [10] Gnawali O, Fonseca R, Jamieson K, et al. Collection Tree Protocol [C]//Culler D E, Liu J, Welsh M, et al. Proceedings of the 7th ACM International Conference on Embedded Networked Sensor Systems(SenSys). Berkeley: ACM,2009. 1-14.
- [11] Guillermo B, François I, Gunnar S, et al. The Hitchhiker's Guide to Successful Wireless Sensor Network Deployments[C]// Abdelzaher T F, Martonosi M, Wolisz A, et al. Proceedings of the 6th ACM International Conference on Embedded Networked Sensor Systems(SenSys). Raleigh: ACM,2008. 43-56.
- [12] Vicaire P, He T, Cao Q, et al. Achieving Long-Term Surveillance in VigilNet[J]. ACM Transactions on Sensor Networks(TOSN), 2009,5(1):1-39.



樊 镭(1984-),男,硕士研究生,研究方向为无线传感器网络中的异常检测和数据挖掘,fnchnw@gmail.com;



龚闻天(1983-),男,硕士研究生,研究方向为无线传感网络路由协议和拓扑控制技术,wentiang@gmail.com;



施晓秋(1964-),女,教授,主要研究方向为网络与分布式计算、无线传感器网络,sxq@wzu.edu.cn。