

无线传感网络中数据融合体系架构的研究^{*})

程宏兵^{1,2} 黄晓¹ 杨庚¹

(南京邮电大学计算机系 南京 210003)¹ (江苏广播电视大学信工学院 南京 210017)²

摘要 数据融合技术是无线传感器网络(WSN)的一个关键技术,目的是减少传感节点间的传输量,降低整个网络中的能量消耗和数据冲突,进而优化 WSN 的整体性能。文章针对 WSN 中数据融合处理形式的多元性,在讨论了广义的数据融合,分析现有的 WSN 数据融合方法的基础上提出了一种解决 WSN(自组织和网内处理)中数据融合多元性问题的新的开放式的体系架构,并给出了具体实现方案。

关键词 无线传感器网络,数据融合,自组织,网内处理,融合体系结构

Research of Data Fusion System Architecture in Wireless Sensor Network

CHENG Hong-Bing^{1,2} HUANG Xiao¹ YANG Geng¹

(Department of Computer Science & Technology, Nanjing University of Post Telecommunications, Nanjing 210003)¹

(School of Information Engineering, Jiangsu Radio & Tv University, Nanjing 210017)²

Abstract Data fusion is a key technology in WSN, it aims to reduce the number of transmissions of sensor nodes, and hence minimize the overall power consumption and the probability of collision in the network, thus to improve the whole performance of WSN. First the paper discusses extent ranging data fusion. Then based on some approach of data fusion in WSN, the paper gives a new open system architecture of data fusion and the architecture's implementation plan, which can apply to WSN(including self-organization and in-network processing.)

Keywords Wireless sensor network, Data fusion, Self-organization, In-network processing, Fusion system architecture

1 引言

计算机网络自产生之日起,尤其是 20 世纪 90 年代初的迅猛发展,使人们的生活发生了极大的变化。具有无线通信、数据采集和处理、协同合作等功能的无线传感器节点协同组织起来形成的 WSN 是一类无线 ad hoc 网络^[1]。WSN 是当前信息技术的前沿之一。它是由大量无处不在的、具有无线

通信与计算能力的微小传感器节点构成的自组织分布式网络系统,在整个网络系统中,大量的传感器节点收集、处理,并且交换来自于外界环境的数据,最终传输到外部基站。该系统是能根据环境自主完成指定任务的“智能”系统,它集成了传感器技术、微机电系统(MEMS)技术、无线通信技术和分布式信息处理技术。图 1 给出了目前一种公认的 WSN 体系结构。

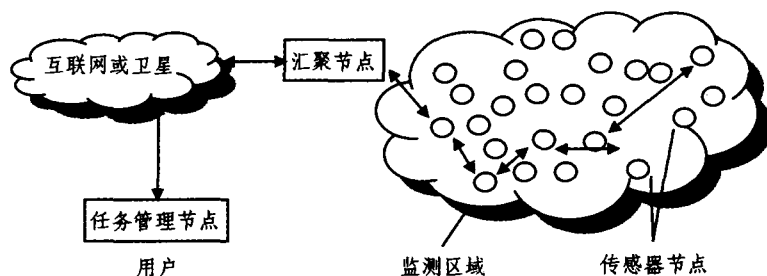


图 1 WSN 网络体系结构

通信技术日新月异。近年来,无线通信更是大行其道,各种无线通信应用层出不穷。无线通信和有线通信相比较,无需对网络通信的物理通道进行配置,具有很大的灵活性,并节约了网络设施的成本,这使得它比有线通信具有更加光明的前景。全球无线通信应用的规模,将很快超过有线网络应用的规模。

在设计无线传感器网络时,基本的问题是克服能量与带宽的限制^[2]。数据融合作为一个基本的方法,其目的是减少传感节点间的传输量,达到降低整个网络中的能量消耗和数据冲突从而优化 WSN 的整体性能。

本文在第 2 节对国内外在各个领域提出的广义数据融合方法(算法)进行综合的分析和比较,得出各方法(算法)的性

^{*} 国家自然科学基金资助项目(编号:70271050),江苏省自然科学基金资助项目(编号:BK2005038)。程宏兵 博士研究生,主要研究方向为信息安全、网格计算、计算机网络;黄晓 博士研究生,主要研究方向为信息安全、网格计算、计算机网络;杨庚 教授,博士生导师,主要从事计算机通信与网络、并行与分布式计算、信息安全方面的教学与研究。

能以及优缺点;第3节讨论了WSN中自组织和网内处理两方面的数据融合问题;第4节针对WSN,提出了符合自组织和网内处理两类数据融合的开放式体系结构,并给出其具体实现方案;最后是文章的总结。

2 广义数据融合

数据处理的应用与研究已经从大型科学数据采集、计算和军事领域的目标识别、跟踪发展到现在网络和个人数据信息的提取和融合。数据信息融合的基本目标是通过融合方法对来自不同来源、不同模式、不同媒质、不同时间、不同地点、不同表现形式的数据进行融合,最终得到被感知对象更加精确、精练的描述。

2.1 数据信息融合后的形式

首先定义一个广义的数据信息集 Ψ :

$$\Psi ::= \{D_1, D_2, \dots, D_n\}$$

$$D_i ::= \{A_1, A_2, \dots, A_m\} \quad i \in [1, \dots, n]$$

$$C_j ::= \{C_1, C_2, \dots, C_k\} \quad j \in [1, \dots, m]$$

$$D_i \cap D_j = \Phi \wedge D_i \cup D_j \neq \Phi \quad (D_1, D_2 \in \Psi)$$

$$A_i \cap A_j = \Phi \quad (A_1, A_2 \in D_k)$$

$$C_i \cap C_j = \Phi \quad (C_1, C_2 \in C_k)$$

其中, Ψ 为经过数据融合后的某一数据信息集, D_i 为第 i 个数据信息源, 可以是来自传感器节点的信息, 也可以是其它例如图档资料、Web 信息等。 A_j 为数据信息源 D_i 中的主题 j 的数据信息集合, C_j 为主题的某一媒质的特征信息。

对于数据信息集 Ψ 来说, 其结果与数据融合的关联方法有重要关系。而关联的关键就是在各个不同级别的融合处理中找到一个等价划分: $g(a_1, a_2, \dots, a_n) = \{S_1, S_2, \dots, S_n\}$, S_i ($i \in [1, m]$) 是关于某一特征或某一主题的信息集。目前, 数据的关联方法大体上分为智能化方法和非智能化方法, 本文将在 2.2 节对它们进行具体的论述。

2.2 关联方法的研究与发展

数据信息的融合的关键是对多源数据信息进行关联, 然后再采用相应的方法对各个聚类中的数据进行无缝集成。国内外提出的关联方法按数学方法可以分为 4 类即概率论、证据理论、模糊逻辑和神经网络。其中的概率论在关联应用中的研究趋于完善, 模糊逻辑和神经网络处于起步阶段。

在军事的应用领域中, 数据关联的研究经历了从单目标跟踪的最近临法、概率数据关联法等到多目标跟踪的联合概率数据关联和多假设理论等。

最早的最近临法(NN)主要用在位置特性的关联上, 对传感器的跟踪中对目标定位上, 把跟踪门限内离目标中心最近的测量作为目标的更新状态, 该方法简单, 但同时存在较大的误差, 相关研究表明^[3], NN 算法只适用于目标较少且非常稀疏的情况下。

概率数据关联(PDA)是由 Bar-Shallom 和 Tse 提出的, 是一种全邻的后验方法, 由落入门限内的测量的加权值来更新航迹, 权值为每个测量属于正确测量的概率, 这种方法是基于航迹已知的情况, 因此有一个严重的缺陷即没有提供航迹起始和航迹终止的方法。

Bar-Shallom 提出了基于 PDA 的联合数据关联方法(JPDA), 该方法适用于对目标的跟踪问题, 由于 JPDAF 算法对所有可能的目标关联解进行搜索, 并在此基础上计算出最佳关联概率 β , JPDAF 中对关联解的搜索是一个求组合数问题, 其搜索过程的计算量随目标和回波数增长呈指数增长趋势, 因此, 该算法的计算量在目标和回波密集时呈“组合爆炸”趋势, 在实际中难以广泛应用。

有的研究者提出了一种提高计算速度的 Viterbi 算法, 用于对单传感器的测量, 其采用一种方格图的方法, 通过找出从开始到结束的最短路径的方法来实现关联, 此算法可以进行并行计算, 从而提高计算速度。在目标的跟踪和文本数据信息的融合中该方法得到了较多的应用。

由于收集的信息中存在不确定因素以及各信息源存在不同的可信度, 因而需要引进模糊逻辑、证据理论和神经网络等智能化方法来进行数据信息融合的数据关联研究, 这些方法都是采用一种非直接的关联方法。前面分析的算法研究大多集中在军事领域, 针对传感器数据融合中目标跟踪和识别进行研究的, 主要是对运动状态的数据信息的关联, 然后加以融合。从本质上讲, 数据形式不同, 其融合关联方法也不一样。随着应用范围的扩展, 研究的问题不断的深入, 各个信息处理系统不再处理单一的信息, 关于异质数据信息的融合问题已成为研究的焦点^[4]。

3 WSN 的数据融合

无线传感网中传感节点之间、传感节点与控制中心之间必须进行通信, 然后对通信结果做进一步的处理。目前所有的通信都依赖于射频(RF), 因此通信时, 必须考虑的是在射频广播通信中有限带宽这一局限问题。传感网络中数据包的建立和维护, 以及从监测点捕获的大量信号, 将对射频媒介的通信能力产生巨大的挑战。

无线传感网络中另外一个重大的局限在于传感节点的低能耗需求, 传感节点一般携带有限的、不可替换的能源, 通过更换数量众多的传感节点的电池来维护传感节点能量的做法显然是不可取的。因此, 当传统的网络致力于获得高质量网络服务时, 传感网络协议则更多关注传感节点能量的保持方面。对于传感网络中能量的耗费主要有两个方面: 计算和通信, 相关实验表明^[5]在 100 米的通信线路上传输 1kb 的数据包和 CPU 执行 3Mb 的指令所消耗的能量几乎是相同的, 因此, 传感节点应注重于本地的数据处理而减少远距离的数据传输, 从而可以减少通信所带来的能量耗费巨大的负担。

数据融合技术在传感网络中已经成为一个非常有用的方法。基于自组织机制和网内处理的数据融合技术可以减少 WSN 建立和维护的代价、消除数据冗余、减少传输次数。于是可以减轻网络带宽的负担及成千上万的数据包处理所带来的能量损耗。下面就网络自组织性和网内处理两个方面对 WSN 中数据融合技术进行讨论。

3.1 自组织性

传感器的工作原理是传感器之间不需要进行任何的连接就可以工作, 传感节点的优势主要是其能快速随机地配置到被监测区域, 这也意味着这些传感设备的无线连接不能配置成一种预定的拓扑结构, 操作时, 传感器很难甚至不能被访问, 因此, 这样的网络需要有自动操作的能力。并且, 由于使用时间的因素, 某些传感器由于能量耗尽而又无法替换, 所以, 各传感器之间必须获悉彼此的情况, 并组成相应的网络, 这种网络称为自组织网络。

目前主要从两个方面来研究无线传感器网络的自组织性, 一个是数据扩散方法, 另一个是网络的拓扑结构, 这两个方面是决定无线传感器网络通信的关键因素, 下面分别进行论述。

3.1.1 数据扩散方法 在 WSN 的数据扩散方法中, 以数据为中心的方法最令人关注。Krishnamachari 等比较了以数据为中心和以地址为中心的数据扩散方法, 其区别在于数据从源点到汇点的传送方式: 在以地址为中心的数据扩散协

议中,每个源点各自沿最短路径传送数据到汇点(端-端),在以数据为中心数据扩散协议中,源点传送数据到汇点所经过的路由节点能获悉数据的内容并在多个数据包之间进行聚合。

以数据为中心的数据扩散方法是在请求通信时,传感节点发布已定义属性的数据请求,当其它节点暂时没有该类数据时,将不会响应,但会记下这样的请求,当后来有符合该属性的数据时,请求数据节点和拥有该数据节点之间将会进行通信,完成数据的传送。以数据为中心的数据扩散方法的优点在于,当需求的数据不存在时将不会发生数据交换,但节点必须要周期性地更新请求且请求方和发送方节点之间必须保持路由信息。Directed Diffusion 方法是传感网络中重要的以数据为中心的数据扩散方法。

The SPIN-X(Sensor Protocols for Information via Negotiation)^[6],是一种基于协商的信息扩散协议,也是一类三次握手协议,其包含两个主要的创新-协商和资源适配。在 SPIN-1 协议中,节点之间利用‘元数据’进行数据传送之前要进行协商。‘元数据’是对传感网中采集到的数据简洁而完整的描述。可以达到消除数据冗余节约传感器能量的目的。SPIN-2 又称 SPIN-1 with a low-energy threshold,是在 SPIN-1 的基础上加入了能量的考虑,对于一个 SPIN-2 节点,当它的能量是很充足的时候,它的表现和 SPIN-1 节点是一样的,通过三次握手协议进行通信。但是当某一节点的能量降到某一数值以下时,它将不再加入三次握手的过程中,此时不再理会外部的一些消息。

3.1.2 WSN 拓扑结构 WSN 的拓扑结构主要分为聚合树和 clustering 两类。通过数据融合,两种结构都能达到最小化传输成本以获得最大化 WSN 生命周期的目的。目前典型的拓扑控制可以分为以下几个方面:功率控制方面有基于节点度的算法和基于临近图的算法。层次型拓扑结构控制方面有 LEACH 算法、GAF 算法、GAF 改进算法及 TopDisc 算法;启发机制方面有 STEM 算法和 ASCENT 算法。

3.1.3 时钟同步 在 WSN 中为了节约能量,尽最大可能使传感节点保持低能耗状态是十分必要的,时钟同步机制在节约传感网络能量方面具有重要的影响。目前 WSN 的网络时间同步机制有 RBS 同步机制、TPSN 时间同步协议、mini sync 和 tiny sync 同步算法、DMTS 同步机制和 LTS 机制等。

3.2 网内处理

基于上述数据扩散方法和网络拓扑结构控制算法,可以为 WSN 构建某一拓扑结构并设定路由,每个传感器在网络中同时扮演感知外界环境、路由其它传感器的作用。接下来的一个问题是 WSN 中一些传感器所产生的数据可能足以使网络发生拥塞,并且产生的这些数据大多数对终端用户来说都是毫无用处的。实际上,可以在这些数据传输之前先进行一定的处理,这就是网内处理。网内处理因为可以消除大量的冗余、无效数据且部分传感器的结果数据可以进行合并、融合,因此可以极大改善 WSN 的整体性能。下面将讨论网内处理的基于聚合函数的数据融合问题。

3.2.1 基本聚合函数 目前,常见的一些基本聚合函数有^[7]: Duplicate suppression, Average, Sum, Minima, Median, Maxima, Percentiles and Count 等,对于它们的性能主要从四个方面(Duplicate sensitive, Exemplary/Summary, Monotonic, Partial State)进行度量。这四个方面的因素对 WSN 的数据处理非常重要。

3.2.2 复杂聚合函数 虽然基本聚合函数适用很大范围的数据应用问题,但对于传感器网络的数据融合问题需要

更加复杂的聚合函数来克服网络能量和资源的局限性,相关研究表明,结合 WSN 的自身特点设计其数据聚合方式时须考虑传感器网络的资源局限性。

DADMA^[8]是一种类数据库的复杂聚合方法,该方法中,每个传感器处有一个本地表单而在汇点有一 SNDV(sensor network database view)。用户在 SNDV 中获取其所需要的数据并用相应的聚合函数来消除数据冗余。

Cam H 等提出了一种基于群的 WSN 协议—Energy-efficient and Secure Pattern-based Data Aggregation(ESPD)^[9]。在 ESPD 中,传感节点不是把整个数据进行传送,而是发送相应模式的代码给群中的主机来进行数据聚合。另外,传感数据以加密形式传送到基站并且在传输路径上保持加密形式,这样就确保了 WSN 数据通信的安全。

以上 2,3 节讨论的各种数据信息融合方法,都是针对各自的应用领域的特定体系框架而设计与实现的,其可移植性不好。就 WSN 中的不同层次的数据融合问题,结合我们以前的研究,在下一节中我们提出了一种适用于 WSN 中各层的开放式数据融合框架,并给出其具体实现方案。

4 WSN 数据融合框架及实现方案

基于以上的研究,论文提出了一种适合于无线传感器网络中数据融合的开放式框架,可以满足无线传感器网络中各个不同层次的数据融合情况。这种基于开放的 WSN 中数据融合框架具有如下特点:(1)传感器数据源具有独立性;(2)数据融合具有并行性、分布性;(3)数据信息的融合具有近似实时性;(4)软件的模块化;(5)功能框架的可扩展性。图 2 给出了 WSN 中数据融合系统的体系结构框图,此体系结构是基于技术驱动模型和数据驱动模型的组合。

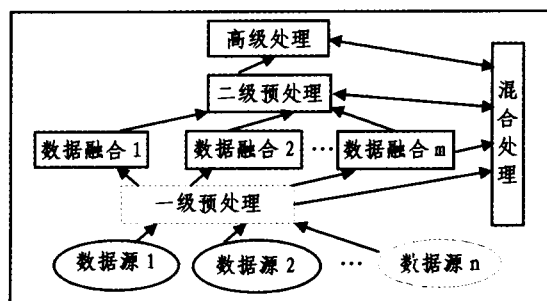


图 2 WSN 中数据融合的体系框图

4.1 模块功能介绍

(1)一级预处理:对传感器网络中各层次数据信息进行初级处理,例如数据的提取、量化、格式转换,处理误差以及按照相应的要求向上作多路传输等。

(2)数据融合集模块:将预处理传来的数据进行分类处理,按照相应的规则分别调用不同的融合模块(具有相应的融合方法,适合不同的数据融合情况)进行处理,从而得出各类比较完整、准确的数据信息,然后传给混合处理或二级预处理等。

(3)二级预处理:是将数据融合集模块处理的分类信息按照更高一级的处理规则进行处理,得出更加精练和更加符合要求的数据信息。此时的结果可以传给混合处理或高级处理。

(下转第 106 页)

$$R(b,d)=\frac{C_{b,d}+C_{d,b}}{2}=\frac{0+0}{2}=0$$

$$C_{b,i}=\text{confidence}(b \Rightarrow i)=P(i|b) \\ =\frac{\text{support_count}(b \cup i)}{\text{support_count}(b)}=\frac{1}{1}=1$$

$$C_{i,b}=\text{confidence}(i \Rightarrow b)=P(b|i) \\ =\frac{\text{support_count}(i \cup b)}{\text{support_count}(i)}=\frac{1}{1}=0$$

$$R(b,i)=\frac{C_{b,i}+C_{i,b}}{2}=\frac{1+1}{2}=1$$

同理可得: $R(b,g)=0, R(e,d)=0, R(e,g)=0, R(e,i)=1, R(i,d)=1, R(i,g)=1$ 。那么,

$$d(p_1, p_5) = \frac{R(b,d)+R(b,g)+R(b,i)+R(e,d)}{3 \times 3} + \frac{R(e,g)+R(e,i)+R(i,d)+R(i,g)+R(i,i)}{3 \times 3} \\ = \frac{0+0+1+0+0+1+1+1}{9} \\ = 0.555$$

可见,用本文提出的方法度量 Web 访问路径的差异性比文[2]的方法更为客观(因 Web 用户访问路径 p_1, p_5 浏览了相同的网页,应具有较高的差异性值)。

结论 通过本文的方法,可以有效地计算出所有 Web 用户访问路径彼此的差异性值,得到 Web 用户访问路径差异性矩阵,为进一步的 Web 用户访问路径聚类操作提供可靠的数据。

(上接第 62 页)

(4)混合处理:将数据融合集处理的结果与二级预处理得到的结果进行综合处理,其处理结果可以传给二级预处理或高级处理。

(5)高级处理:将前面各相应模块传来的数据处理结果进行综合处理,通过事先制定的相应规则及逻辑推理方式得出所需要的最终数据融合结果,并输出之。

4.2 体系框架的实现

为了验证我们所提出的 WSN 中数据融合体系框架的有效性,证明其能够对 WSN 中各层次的数据融合按照不同的要求进行快速开发,我们采用一种主体框架与各个功能组件相结合的形式实现这一框架。图 3 是各个功能模块通过组件技术实现的数据融合体系框图。

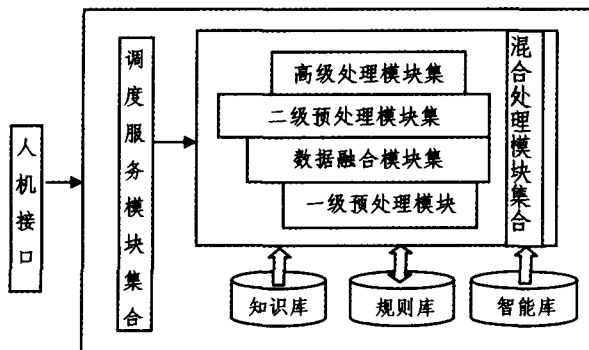


图3 基于组件技术的处理实现框架

该数据融合架构由于采用组件的技术来实现,因此对于涉及到的模块可以动态地卸载和加载,具有很大的灵活性。在我们进行的 WSN 网络基于自组织和网内处理的数据融合仿真实验中,该框架都显示了很强的适应性。对于不同的融合要求只需对框架中的相应模块进行调整,使其满足具体的

将本文的方法用于实际 Web 网站中,再施以恰当的聚类操作,即可在网站运营中,根据不同用户的网页访问情况,对其进行个性化推荐、页面结构动态调整、页面集预存储等操作。具体的作法是:先对已有的 Web 用户访问路径进行聚类,将具有相似浏览兴趣的 Web 用户访问路径聚集成一类;然后在当前用户的浏览过程中,实时地将其访问路径与已有聚类相比较,并将其聚类到最接近的某一类别中,再根据这一类别的浏览兴趣或浏览目的对当前用户进行相应的个性化操作。

因此,作者进一步的工作就是研究好的聚类算法,在本文得到的 Web 用户访问路径差异性矩阵基础上对 Web 用户访问路径实施有效的聚类操作。

参考文献

- 1 Buchner A G, et al. Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining. ACM SIGKDD Record, 1998, 27(4): 54~61
- 2 Xu Baowen, Weifeng, Song William, et al. Application of Data Mining in Web Pre-Fetching. IEEE Multimedia Software Engineering. In: Proceedings. International Symposium on, 2000. 372~377
- 3 Nasraoui O, Krishnapuram R, Joshi A. Relational Clustering Based on a New Robust Estimator with Application to Web Mining. In: Proc. Intl. Conf. North American Fuzzy Info. Proc. Society (NAFIPS 99), 1999. 596~607
- 4 王晔,李德毅. 自适应 Web 站点的访问数据聚类方法[A]. 中国人工智能进展[C]. 北京:清华大学出版社, 2001. 402~406
- 5 陈恩红,徐涌,王煦法. Web 使用挖掘:从 Web 数据中发现用户使用模式. 计算机科学, 2001, 28(5): 85~88

应用要求。

结束语 数据融合技术是无线传感网络中的一个关键技术,对于其中数据融合情况由于针对不同的特性和不同的分类方法存在很大差异,因此 WSN 中数据融合的处理方式也不一样,没有一个统一的处理模式。本文提出了一种基于组件和智能技术的 WSN 数据融合架构,仿真实验表明,这种系统框架可以按照 WSN 中不同的数据融合进行动态灵活的调整,显示了很强的适用性,可以作为 WSN 中数据融合的统一架构。

参考文献

- 1 Pottie G J, Kaiser W J. Embedding the internet: wireless integrated network sensors [C]. Communications of the ACM, 2000, 43(5): 51~58
- 2 Kalpakis K, Dasgupta K, Namjoshi P. Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks [M]. [TR CS-02-12]. Aug. 2002
- 3 Zhou B, KBOSE N. Multitarget Tracking in clutter: Fast Algorithms for Data Association [J]. IEEE, Trans On Aerospace and Electronic System, 2003, 29(2)
- 4 Germain M, Voorons M, Boucher J M, et al. Fuzzy statistical classification method for multiband image fusion. Information Fusion [C]. In: Proc. of the Fifth Intl. Conf. 2002
- 5 Ghiasi S, Srivastava A, Yang X, Sarrafzadeh M. Optimal Energy Aware Clustering in Sensor Networks [A]. Sensors Magazine, MDPI, Issue, January. 2002. 258~269
- 6 Heinzelman W, Kulik J, Balakrishnan H. Adaptive Protocols for Information Dissemination in Wireless Sensor Networks [J]. In: Proc. of 5th ACM/IEEE Mobicom Conference, 1999
- 7 Madden S, Franklin M J, Hellerstein J M, Hong W. TAG: A Tiny Aggregation Service for ad-hoc Sensor Networks [M]. OSDI 2002
- 8 Cayirci E. Data Aggregation and Dilution by Modulus Addressing in Wireless Sensor Networks [A]. Communications Letters, IEEE, 2003, 7(18): 355~357
- 9 Cam H, Ozdemir S, Nair P, Muthuvinashiappan D. ESPDA: Energy-efficient and Secure Pattern-based Data Aggregation for wireless sensor networks [J]. Sensors, 2003. In: Proc. of IEEE, 2003, 1(22-24): 732~736