

无线传感器网络数据收集研究进展

解文斌 鲜 明 包卫东 陈永光

(国防科技大学电子科学与工程学院 长沙 410073)

摘 要 在无线传感器网络中,数据收集技术至关重要。在归纳无线传感器网络数据收集技术研究进展的基础上,分析了近年来该领域具有代表性的数据收集算法,以网络结构、流量优化和移动性为依据分为三大类,并分别指出了这些算法的特点和适用情况,最后总结了数据收集算法未来的研究策略和发展趋势。

关键词 无线传感器网络,数据收集,网络结构,流量优化,移动性

Overview of Data Gathering in Wireless Sensor Networks

XIE Wen-bin XIAN Ming BAO Wei-dong CHEN Yong-guang

(School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract Data gathering technology is crucial in wireless sensor networks. Based on the summary of the progress in data gathering technology, the recent representative data gathering algorithms in this field are analyzed. Then these algorithms are classified to 3 categories based on network structure, flow optimization and mobility, and their characteristics and application areas are compared respectively. Finally, the research strategies and trends of data gathering algorithms are summarized.

Keywords Wireless sensor network, Data gathering, Network structure, Flow optimization, Mobility

无线传感器网络(WSN)由具有感知和通信能力的传感器节点组成。传感器网络技术是支持普适计算和未来感知系统的重要技术基础,对传感器网络相关技术的研究已经成为计算机网络、通信和微电子领域中的一个热点。随着技术的成熟和市场的推广,传感器网络技术将应用于越来越多的领域,例如战场监视、医疗护理、污染监测和目标跟踪等。在这些应用中,基本的操作是数据收集。传感器网络是以数据为中心的网络,灵活高效的网络数据收集机制是一个必须解决的关键问题。

与传统的有线网络和无线网络相比,传感器网络的数据收集面临更多的技术难题:(1)能效问题。在传统的有线网络和无线网络中,首要目的是提高服务质量和高效带宽利用,其次才考虑节能问题。而传感器网络里由于节点的电池能量有限,首要的设计目标是能源的高效使用。尽管目前已经从硬件设计、节能路由协议到应用层提出了大量的建议并得以实现,但是在当前技术条件下传感器节点的寿命还是非常有限。(2)数据相关性。由于节点密集部署的特点,邻近传感器产生的数据通常是冗余和高度相关的。所有传感器直接向基站传输数据的做法是效率低下的。此外,大型传感器网络产生的数据量通常过于巨大,给无线传输带来巨大压力,基站也难以处理。因此,必须考虑利用感知数据的时间和空间的相关性,减少冗余数据,从而减轻无线链路和基站的开销。(3)分布式自组织。传感器节点的计算和存储能力有限,无法进行复杂计算,更无法存储全局网络拓扑。因此,数据收集机制必须是全分布式的,且要求具有自组织能力。(4)交互式数据收集。在某一特定时段,观测者可能只对某一区域的某些类型的传感器收集的数据感兴趣,即期望的读数集合较少,没有必要激

活大量传感器节点的数据收集过程。(5)自适应性。传感器网络的拓扑可能由于多种原因(节点失效、链路干扰等)而发生变化。这要求数据收集机制能够适应这些变化,具有动态的自适应性。以上的这些要求使得一些常规的数据收集和处理机制不能很好地发挥作用。

目前已经提出了一些许多用于解决传感器网络数据收集的算法和系统。但是根据应用背景的不同,它们在网络结构、能量需求、时空复杂性等方面有所不同。本文综述了传感器网络数据收集的相关技术和分类,总结了近年来典型的算法。

1 传感器网络数据收集问题

1.1 相关概念

(1)数据收集:定义为从多个传感器系统地收集环境参数的感知数据,最终传输到基站处理的过程。本文中用 sink 表示基站。

(2)数据聚合:定义为网内融合多个传感器的数据以去除冗余数据,为基站提供融合信息的过程。数据聚合试图从传感器收集最关键的数据,使其以一种数据延迟最小的能量高效的方式发送到 sink。显然,数据聚合是数据收集的特例。聚合是减少数据容量的一种有效方式,但对于需要详细的感知数据的应用(如目标跟踪)是相当粗略的。

数据聚合是一种节能技术,试图减少通信的数据量,由中间节点收集局部数据,仅向 sink 转发一个聚合操作的结果,例如 min 和 max。

一般数据聚合问题是,在 n 个节点的网络($m < n$)里给定 m 个源和一个 sink,寻找一个包括所有源的最小加权子图。这是一个著名的 NP-完全问题,已知为 Steiner Tree 问题

解文斌 博士生,主要研究领域为传感器网络、信息安全和电子对抗;**鲜 明** 博士,副教授,主要研究方向为信息安全与电子对抗;**包卫东** 博士后,副教授,主要研究方向为指挥自动化与传感器网络;**陈永光** 总工程师,研究员,博士生导师,主要研究方向为电子信息装备试验技术、电子对抗等。

(STP)。

1.2 传感器网络数据收集算法的性能评价

传感器网络数据收集算法的性能直接影响其可用性。下面定性地讨论几个常用的评价标准。

(1)能效。首要的评价指标就是能效,一般情况下都使用网络生命期衡量算法的能效高低。在一个理想的数据收集机制里,在每一个数据收集轮次里传感器节点消耗同样数量的能量。网络生命期定义为直到 $\alpha\%$ 的传感器节点失效时数据收集的轮次数,其中 α 由设计者决定。例如,在多数文献里认为每一个节点都很重要,网络生命期就定义为直到第一个节点失效时数据收集的轮次数。

(2)数据准确性。数据准确性的定义取决于特定应用的要求。例如,在一般的环境监测里对数据准确性的要求不是很高,而在目标跟踪里就要求较高的数据精度。

(3)执行时间。执行时间定义为包含数据传输、路由和数据聚合或压缩的时间。它可以用 sink 接收报文和源节点产生数据之间的延迟时间来表示。不同的应用对数据新鲜性有不同的要求,因此对数据收集算法的延迟也有不同的要求。

(4)可扩展性,即伸缩性。

(5)鲁棒性,即容错性。

(6)网络吞吐量。网络吞吐量定义为 sink 接收到的最大数据流量。

上述性能指标不仅是评价 WSN 数据收集算法的标准,也是其设计和实现的优化目标。值得注意的是,以上性能指标并不完全兼容,没有一个算法可以在所有方面达到最优。

2 基于网络结构的数据收集协议和算法

WSN 的网络结构是影响数据收集协议的一个重要因素。针对多对一的平坦型和层次型通信结构,Duarte-Melo 等人^[1]研究了不同通信结构里传感器网络数据收集的传输容量与能耗问题。在网络结构、容量和能耗之间发现一些有趣的关系:如果网络区域很大,平坦型网络消耗更少的能量;如果网络规模小,层次型网络消耗更少的能量。

2.1 平坦型(Flat)网络

在平坦型网络里,包含大量的静态传感器节点和一个 sink,所有传感器节点拥有同样的能量,具有相同的角色,都能够与 sink 通信。报文通过一跳或多跳转发到 sink,即所有数据流量都流向 sink。因此,靠近 sink 的节点比处于网络边缘的节点消耗更多的能量。这种网络结构只适合于小规模的网络。

(1)两阶段 pull 扩散

定向扩散(Directed Diffusion)^[2]是一种有代表性的两阶段 pull 扩散方法。sink 节点周期地广播“兴趣”消息,告诉网络中的节点它需要什么数据。节点通过建立“梯度”来转发“兴趣”给它的邻居节点,梯度指向“兴趣”消息的来源方向。中间节点能够缓存和变换数据。定向扩散使用局部数据驱动的规则实现数据收集任务,比多播扩散的能效高很多,但不适合要求连续向 sink 发送数据的应用。定向扩散的性能受到源和目标节点位置、网络拓扑等因素的影响,当源相互靠近且远离 sink 的时候能够节约大量的能量。

(2)一阶段 pull 扩散

对于包含多个源和多个 sink 节点的网络结构,两阶段 pull 扩散将花费大量的开销用于建立传输路径。Krishnamachari 等人^[3]提出了一阶 pull 扩散机制,跳过定向扩散的洪泛

过程。在一阶 pull 扩散里,sinks 发送的“兴趣”消息通过网络建立的梯度传播。然而,源并不传播探测消息。源只向延迟最短的 sink 传输数据。因此,反向路由(从源到 sink)具有最短延迟。去除探测消息减少了控制开销,从而节约了传感器的能量。为了区分不同的 sink,每一个兴趣携带一个流量 ID。

(3)push 扩散

在 push 扩散里源节点是主动的:当它们探测到事件时,就向网络扩散事件消息。而 sink 节点是被动的:它们响应到事件消息后,若对此事件感兴趣,就向源反馈一个强化消息,“加强”梯度最大的路径。通过协商的传感器信息协议 (SPIN)^[4]可以归为基于 push 的扩散协议。SPIN 的两个主要特征是协商和资源自适应。为了成功实现数据协商,传感器节点用元数据来简洁地描述观测数据。

这种机制适合于 sink 节点较多、源节点较少的网络结构。

2.2 层次型(Hierarchical)网络

层次型结构通过在网络里增加少数高能量的节点或选举某些节点作为簇头,把整个网络自组织成不同的层次,增强了扩展性。

(1)簇结构的网络

在能量受限的传感器网络里,传感器节点直接向 sink 传输数据的效率较低。传感器节点先向簇头发送数据,经簇头聚合后向 sink 传输简洁的数据摘要,可以节约大量的能量。近年来提出的较有代表性的协议是 LEACH^[5], HEED^[6], TEEN^[7] 和 CLUDDA^[8]。

LEACH^[5]的基本思想是通过周期性等概率地随机选择簇头,将整个网络的能量负载平均分配到每个传感器节点,在簇头执行数据聚合,从而达到降低网络能耗的目的。尽管 LEACH 增加了系统生命期和数据的准确性,但要求簇头节点具有较大的通信能力,扩展性差,不适合大规模的网络;簇头节点的能量消耗也很快;频繁地选择簇头大大增加了网络能耗。

HEED^[6]的主要目标是通过高效成簇,将能耗均匀分布到整个网络从而最大化网络生命期。HEED 的簇头选择主要依据主、次两个参数。HEED 的主要改进是:在簇头选择中考虑了节点的剩余能量,并以主从关系引入了多个约束条件作用于簇头的选择过程。HEED 在簇头选择标准以及簇头竞争机制上都与 LEACH 不同。实验结果表明,HEED 分簇速度更快,能产生更加分布均匀的簇头、更合理的网络拓扑。

TEEN^[7]采用类似 LEACH 的成簇算法,只是在数据路由阶段使用不同的策略。具体做法是在协议中设置了硬、软两个阈值,以减少发送数据的次数。在每轮簇头选举的时候将两个阈值广播出去。通过调节两个阈值的大小,可以在精度要求和系统能耗之间取得合理的平衡,减少网络通信量。但 TEEN 不适合需周期性上报数据的应用。

CLUDDA^[8]是一种结合了成簇和定向扩散的混合方法。主要分为两个阶段:兴趣传播和数据传播。兴趣消息里包含查询定义,描述需要对数据执行的操作。在兴趣传播阶段,只有簇头和 sink 执行兴趣分发任务。普通节点保持沉默,直到它能提供查询请求的服务。在簇头缓存查询消息,列出需要聚合的不同数据成分。在数据传播阶段执行分层聚合任务,聚合点是动态的,随源节点位置的变化而变化,尽可能在靠近

源节点的地方执行数据聚合任务。任何具有查询定义的簇头或 sink 节点都可以选为聚合点。

(2)链式的网络

在簇结构的网络里,如果簇头与传感器节点相距较远,通信开销就很大。如果只传输到最近的邻居,就能进一步改进能效。

PEGASIS^[9]是在 LEACH 基础上建立的一种链式的数据收集协议。该协议假设所有节点都静止,通过利用贪心算法,能够以一种中心化的方式形成一条相邻节点之间距离最短的链。在每一个数据收集轮次,节点从上一个节点接收数据,与自己的数据聚合后,沿贪心链把结果传输到下一节点。最终由类似于簇头的领导节点传输到 sink。该协议通过避免 LEACH 协议频繁选举簇头带来的通信开销以及自身有效的链式数据聚合,节点采用低功率与最近邻节点通信,大大减少了数据传输量,与 LEACH 协议相比能大幅提高网络生存时间。主要缺点是:需要知道所有节点的位置信息;假设所有节点具有相同的功率;链尾节点的数据延迟很大;单链方式容错性差,若链路中间某一节点失效会导致路由失败。

(3)树结构的网络

在基于树结构的网络里,传感器节点被组织成一棵树的结构,非叶子节点可以执行数据聚合,沿树枝路径把数据发送到根节点。在该结构下的数据收集问题一般简化为寻找最小代价生成树的问题。严格说来,这也是一种平坦型网络。

EADAT^[10]是用于构造和维护聚合树的一种能量感知的启发式方法。算法的主要思想是关闭叶子节点的射频以节约能量,只有非叶子节点负责数据聚合与流量转发。节点选择剩余能量更多、更靠近 sink 的节点作为其父节点。只需要每个节点广播一次就能得到一棵根为 sink 的树。该算法的主要优点是剩余能量多的传感器成为非叶子节点的机会更多。

(4)基于网格的网络

在基于网格的聚合机制^[11]里,传感器网络的监测环境被分割为一些网格或区域。在每一个网格或区域里选择一个传感器,赋予数据聚合者的角色,负责聚合网格里所有传感器的数据并向 sink 通报关键信息。网格内的传感器之间不相互通信,直接向本网格的聚合者发送数据。类似于基于簇的数

据聚合,但网格里的聚合者是固定的。基于网格的数据聚合适合于网络拓扑动态变化或事件移动的环境。因为监测高动态事件时,网络内传感器之间的数据交换非常频繁,但是一旦到达聚合者之后,只有最重要的信息才向 sink 发送,从而大大减少网络流量并确保关键信息能够及时发送。

2.3 无结构型网络

对于所有节点向 sink 周期性发送数据的应用,因为网络的流量模式稳定,不需要引入大量的传输结构维护开销,结构化的数据收集协议适合于此类应用。对基于事件驱动在动态情景下的应用,构造和维护传输结构的开销所带来的好处可能超出采用节能数据聚合协议所带来的。

Fan 等人^[12,13]提出了一种无结构数据聚合协议,不需要显式地维护传输结构,同时能够获得高效的数据聚合。由于没有显式消息交换,节点不知道它们应该把数据发送到哪里,也不知道它们应该等待多长时间聚合数据。因此,为了增强空间收敛性,文献[12,13]中在 MAC 层提出了数据感知任播(DAA)协议,不预先构造最优聚合结构和流量模式,而是在向 sink 转发报文的过程中自动建立源节点之间的独立集,由独立集中的节点担任聚合节点;为了增强时间收敛性,在应用层提出了随机等待(RW)技术,由每个源节点在 0 到 τ 之间为报文随机选择一个延迟时间,其中 τ 是最大延迟,最优值取决于事件的规模和报文的传输时间。仿真结果说明,联合 DAA 和 RW 的方法比偶然性聚合的性能好 73%,当聚合函数不完美时比结构化聚合的性能好。

2.4 基于网络结构的数据收集算法比较

采用不同网络结构的各种协议具有不同的特点,适用于不同的数据收集应用。为便于说明,采用表格方式加以说明。从表 1 可以看出,平坦结构(包括树结构)的协议一般适合于小规模网络,延迟与网络规模成正比,大多为 sink 主动传输方式。簇、链和网格结构的协议一般适合于中小型网络,但要求网络中有高能节点作为簇头,一般采用单路径方式,延迟较小。此外,平坦型网络便于寻找最优路由,层次型网络不一定能够找到最优路由。但是,结构化方法一般适用于周期性查询驱动的网络,对节点移动性的支持不好。而无结构的方法适合于事件驱动的应用,但是负载均衡和传输延迟的性能较差。

表 1 基于网络结构的数据收集算法比较

协议	结构类型	适用范围	扩散方式	数据聚合	路径数	节能	负载均衡	传输延迟	传输模式	QoS 支持	移动性
Directed	Flat	Sink 较少,周期性查询驱动的小规模网络	Two-phase pull	✓	多路径	差	差	高	主动	可能	不支持
Diffusion		多源、多 sink,周期性查询驱动的小规模网络	One-phase pull	✓	多路径	一般	差	高	主动	可能	不支持
Krishnamachari 等人[3]	Flat	Sink 较多、源较少的小规模网络	Push	✓	单路径	好	差	一般	被动		支持
SPIN	Cluster	节点能量较高的中小型网络		✓	单路径	差	差	一般	主动		
LEACH	Cluster	节点能量较高的中小型网络		✓	单路径	好	好				不支持
HEED	Cluster	节点能量较高的中小型网络		✓	单路径	很好		低	被动		
TEEN	Cluster	节点能量较高的网络	Pull	✓	单路径	好			主动		不支持
CLUDDA	Chain	节点能量较高的中小型网络		✓	单路径	好	差	高	主动		不支持
PEGASIS	Tree	单 sink,周期性查询驱动小规模网络	Pull	✓	单路径	好	好		主动		不支持
EADAT		网络									
基于网格的聚合机制[11]	Grid	节点能量较高的网络		✓	单路径	较好	差	低		可能	支持
DAA 和 RW	Structure-free	基于事件驱动的动态网络		✓	单路径	较好	差	高			支持

3 基于流量优化的数据收集协议和算法

目前很多研究者把传感器网络建模为一个有向图,利用图论和最优化的相关知识研究传感器网络数据收集问题。假设传感器网络部署在某一区域,传感器的位置已知,而且是固定的。一般用有向图 $G=(V,E)$ 建模,其中 V 是节点集, E 是边(无线链路)的集合。

3.1 最大生命期数据收集

MLDA^[14]是一种用于解决传感器网络最大生命期数据收集问题的多项式时间次优算法,其目标是获得一个最大生命期的数据收集调度。在 MLDA 里,根据一阶射频模型建立了传感器的能量模型,并假设系统完全连接,每一轮执行一次数据收集,且每一个传感器只产生一个数据报文,中间节点能够把来自多个源的输入报文聚合成一个简单的输出报文。在

传感器能量约束和边容量约束下,利用整数规划计算出最优的容许流量网络。MLDA 里提出了一种构造聚合树的启发式方法,在每一个数据收集轮次,根据容许流量生成一棵表示数据收集调度的有向树。数据聚合发生在生成聚合树之后。对任意 $\epsilon > 0$,最终能够以 $(1-\epsilon)$ 的比率获得近似最优解。但是,在最坏情况下 MLDA 算法的运行时间是 $O(n^{15} \log n)$,其中 n 是传感器节点数。因此,MLDA 的可扩展性较差,不适应大规模网络的要求。

MaxConcurrentFlow^[15]算法把最大网络生命期问题建模为一个并发多商品流问题,目标是在给定传感器节点能量约束和流量约束条件下最大化网络生命期。其中,每一件商品代表由一个源节点产生并流向 sink 的感知数据,并为每一个节点赋予一个表示其当前负载(通过该节点的数据量)的权值。通过更新节点的权值,在算法的每一次迭代中计算每一件商品的最短路径,最终形成了一棵根在 sink 的最短路径树。利用一种完全多项式时间近似机制(FPTAS),给出了一种最大并发流(MaxConcurrentFlow)算法。对任意 $\epsilon > 0$,能够以 $(1-2\epsilon)$ 的近似度计算出最优生命期。

RFEC^[16]算法解决由最大生命期数据收集问题退化出的一个受边容量约束的流量优化问题,但不考虑数据聚合,目标是确定是否存在一个满足配额、流量守恒和边容量约束的数据流。与标准流量约束问题的不同之处在于 RFEC 引入了额外的配额约束。给定一个图 $G=(V, E)$ 和一个流量 f ,由 f 引出的剩余图是 $G_f=(V, E_f)$ 。 E_f 里的边 (u, v) 具有剩余容量 $c_f(u, v)=c(u, v)-f(u, v)$ 。RFEC 的基本思想就是在 G_f 里寻找从源到 sink 的最短增量路径 P 。RFEC 可获得整数解,基于流量分解技术可以重构出每一轮在相邻传感器之间传输的报文数量文献。文献[16]还提出了一种基于自稳生成树构造算法和最短路径路由的分布启发式方法,但是不一定能获得最优解,因为它只在原始图而不是剩余图里搜索可能的路径。

3.2 网络相关数据收集

在密集部署的传感器网络里,从空间邻近的传感器收集的数据通常是冗余和相关的,同一传感器相邻时刻的感知数据也是相关的。

Cristescu 等人^[17,18]利用 Slepian-Wolf 编码和显式通信的联合熵编码研究了最小代价的网络相关数据收集问题。当传感器使用源编码策略时,目标是通过联合优化节点的速率分配和传输结构来最小化总的传输代价。在 Slepian-Wolf 编码里,最优编码是给靠近 sink 的节点分配更高的速率,给网络末端的节点分配更低的速率,最优编码复杂,传输最优化简单。在显式通信模型里恰恰相反。文献[17]里研究了最优的 Slepian-Wolf 速率分配机制。在单 sink 数据收集问题里,最短路径树(SPT)对基于 Slepian-Wolf 编码的速率分配都是最优的。在这一机制下,最靠近 sink 的节点按等于其非条件熵的速率编码数据。其它节点按等于各自条件熵的速率编码,条件熵是相对于所有比自己更靠近 sink 的节点的熵。为了克服节点需要全局距离知识的缺点,提出了一个仅利用本地邻居节点集的全分布式近似算法。在显式通信模型里,最优速率和传输结构相互影响,联合优化是 NP-hard 问题。文献[17,18]中提出了一些启发式近似算法,包括最短路径树(SPT)、平衡 SPT/TSP 树、模拟退火、叶子删除(LD)和 SLT 算法。仿真结果表明,这些算法能够大大改善总传输代价。

Rickenbach 等人^[19]研究了不同单输入(简单相关边)编

码策略下传感器网络的相关数据收集问题。MEGA^[19]算法首先构造最短路径树(SPT),然后根据节点间的相关系数重新计算边的权值,得到单输入的最小能量数据收集树,在有向最小生成树(DMST)里的下一跳节点执行数据聚合。MEGA 利用外编码方式,把数据收集问题简化为在一个有向图中寻找一个最小生成树(MST)的问题,能够在网络拓扑和节点相关结构已知的条件下导出一个最小能量数据收集拓扑,时间复杂度为 $O(n^3)$ 。文献[19]还提出了一种利用自编码方式寻找最优数据收集拓扑的 LEGA 算法,能够以 $2(1+\sqrt{2})$ 的比率近似最优解,运行时间为 $O(m+n \log n)$ 。

在文献[17-19]里使用了编码/压缩技术,尽管减少了传输量,但所有传感器都需要参与数据传输。Gupta 等人^[20]给出了一种替代方法,通过减少卷入数据收集过程的节点数,抑制某些节点的数据传输,把传感器网络相关数据收集问题定义为一个选择连通相关支撑集(connected correlation-dominating set,CCDS)问题。首先,通过在普通的数据收集消息里捎带确认(piggybacking)邻居节点的消息,计算相关超边参数,构造描述数据相关结构的相关图。然后,选择少量传感器节点子集(连通相关支撑集),构造一个执行数据收集任务的连通的通信图。网络中其他节点的测量值可以以足够的精度从连通相关支撑集导出。文献[20]提出了一系列计算连通相关支撑集的分布式(基本分布式算法、2-轮分布式算法和握手算法),可以应用于异步通信模式,并能够适应相关图的动态变化,容忍消息丢失的影响。还给出了指数幂次的中心化近似算法及相关的多项式时间启发式方法。仿真结果表明,这些方法能够获得足够小的支撑集,即使部分节点失效也能有效延长网络生命期。

SBR^[21](Self-Based Regression)是 Deligiannakis 等人提出的一种压缩传感器历史数据的算法。基本思想是利用同一传感器多个测量值之间在时间上的相关性和冗余性,在获得高度数据约简的同时捕获测量值的细节信息。SBR 算法的关键是引入了基信号的概念。通过从一系列真实数据中抽取基信号特征,把传感器收集的数据分割成不同区间,计算各段数据的线性回归参数,对收集到的测量值进行分段线性相关编码。它是一个用于描述数据特征的字典,存储在传感器里,传输过程中发送到 sink 节点以重构测量值。仿真实验说明 SBR 的性能超过标准的近似技术,如小波、直方图和离散余弦变换。

3.3 QoS 感知的数据收集

在一些应用中,对服务质量有特殊的要求,如带宽、信息吞吐量、端到端延迟和拥塞控制等。针对不同的应用目标,还存在一个取舍与平衡问题。

Ordonez 等人^[22]研究了在能量受限的传感器网络里的最优信息抽取问题。根据加性白噪声信道下的 Shannon 容量方程和信号衰减,把最优信息抽取问题形式化为一个满足能量约束条件的非线性流量规划问题。通过设置节点的传输功率和流量速率,限制节点发送信息量,从而最大化 sink 节点接收的信息量。在文献[22]里给出了两个模型,即满足能量约束的最大信息收集模型与满足信息约束的最小能耗模型,并证明它们在最优解和约束条件上相互等价,但后一种模型的收敛速度更快。根据优化模型,讨论了最优解如何随能量/信息约束、网络规模、公平性因子的变化而变化,比较了直连启发式和多跳启发式在不同拓扑结构下的性能。

Upadhyayula 等人^[23]主要从能效和延迟两个方面研究了

基于树结构的数据聚合算法。文献[23]中研究的是一种特殊的数据聚合过程,网络中所有节点都是源节点,每一层节点都根据压缩因子 γ 对子节点的数据执行聚合操作,然后向单个sink节点转发。通过限制输入节点的数量,引入 β 约束以平衡树结构和降低延迟。首先,修改了Dijkstra的单源最短路径算法,分别给出了最小生成树(MST)算法及其变体。然后引入节点的数据增长因子($\alpha=1-\gamma$),提出一种结合了MST和SPT的算法COM。树节点使用一种信道分配算法和直序扩谱CDMA(DSSS-CDMA)编码实现无冲突通信调度。实验结果表明,对给定网络密度和中间节点的数据压缩因子 γ ,无论主要目标是最小化延迟或能耗,都可以从中选择一种合适的算法。

Ignacio等人[24]研究了不同网内聚合算法的能效、数据准确性和新鲜性之间的差异。结果显示,对于传感器网络里周期性产生数据的应用,停工时间(在转发数据之前等候数据的时间)扮演重要的角色。对于周期性数据聚合,文献[24]中提出了一种新的级联超时聚合机制。根据节点在数据聚合树中的位置设定超时,节点的超时发生在其父节点的超时之前。源自叶子节点的数据及时到达树结构的下一层进行聚合,产生一个流向sink的“数据波”。通过仔细选择等待聚合与转发数据的时间,可以在保证数据新鲜性和准确性的同时节约大量的能量。主要优点是不需要传感器节点之间时钟同步;主要缺点是该方法不适合非周期性数据生成的情况,以及聚合的数据报文长度与生成的数据报文长度不同的应用。

表 2 基于网络流量的数据收集算法比较

算法	目标	假设条件	约束条件	方法	优点	局限性
MLDA	最大化网络生命期	完全连接,每个传感器每一轮只产生一个数据,中间节点能把多个源的数据聚合为单个数据	传感器能量约束和边容量约束	利用整数线性规划计算出最优容许流量,构造数据收集调度树	能够以 $(1-\epsilon)$ 的比率获得近似最优解	计算复杂,可扩展性差,对流量模型做了过多简化
MaxCorr-currentFlow	最大化网络生命期	建模为并发多商品流问题	传感器节点能量约束和流量约束	Garg-Konemann算法和Dijkstra最短路径树算法	能够以 $(1-2\epsilon)$ 的比率获得近似最优解	计算量大,贴近近似度增加性能下降很快
RFEC	最大化网络生命期	不考虑数据聚合,引入配额约束	配额约束、流量守恒约束、边容量约束	利用剩余图方法搜索最短增量路径	能够找到最优解;启发式方法不需要全局同步	对流量模型做了过多简化
Cristescu等人[17,18]	最小化总传输代价	节点感知数据在空间上相关,服从高斯随机模型	边容量约束,即条件熵约束	基于Slepian-Wolf编码,联合优化节点的速率分配和传输结构	利用空间相关性,优化了传输结构和节点的速率分配,大大减少了通信能耗	计算复杂,高斯模型不一定符合实际
MEGA	最小化总能耗	双向链路,节点能够调整射频信号强度,网络拓扑和相关结构已知	条件编码,单输入(简单相关边)编码策略	在SPT基础上搜索DMST	能够在多项式时间内获得最优解	固定的相关结构
Gupta等人[20]	最小化总能耗	节点感知数据在空间上相关		利用最小二乘法计算相关超边,选择CCDS构造数据传输结构	通过空间采样,减少参与传输的节点数,从而大大减少通信能耗;通过近似约简数据大大减少通信能耗;可根据需要调整近似精度,比聚合方法精度高,适合于长期记录感知数据的应用	计算CCDS较复杂
SBR	约简传感器历史数据	节点感知数据在时间上相关,传感器收集到足够多的数据后才向基站发送		从一系列真实数据中抽取基信号特征,对测量值进行分段线性回归	近似精度,比聚合方法精度高,适合于长期记录感知数据的应用	存储量大
Ordenez等人[22]	最大化总的信息收集量	加性白噪声信道,无干扰,信号随距离的平方衰减	传感器能量约束	非线性规划	以形式化方法研究了网络参数对最优解的影响	计算复杂
Upadhyayula等人[23]	最小化延迟或能耗	所有节点都是源节点,每一层节点都根据压缩因子 γ 对子节点的数据执行聚合操作;无冲突通信调度	β 约束	Dijkstra, MST, COM算法	能够在最小化总能耗的同时获得低延迟;	模型过于理想
Ignacio等人[24]	最小化总能耗	周期性收集感知数据	数据新鲜性和准确性约束	级联超时聚合机制	不需要时钟同步	不适合非周期性数据生成的情况,及报文长度不同的应用
Boulis等人[25]	最小化总能耗	周期性收集感知数据	数据准确性约束	基于门限进行传输决策	数据聚合不依赖于层次型的树结构	融合算法依赖于特殊的聚合函数
Yu等人[26]	最小化总能耗	多源、单sink的聚合树结构已知,非单调能量模型	低延迟约束	动态规划	适用于任何聚合函数,可在线调度	结构固定;引入等待延迟

Boulis等人[25]研究了传感器网络里周期性数据聚合的能量-准确性平衡问题。为了分析周期性聚合应用中的能量/准确性子空间,提出了一种联合设计信号处理算法和网络协议的分布式估计算法。关键思想是一个基于门限的机制,把门限用作一个表征准确性和能耗之间平衡的调整参数。该方法的主要优点是执行数据聚合不依赖于层次型的树结构,每一个节点具有聚合数据的全局信息。主要缺点是融合算法的功能依赖于特殊的聚合函数。

Yu等人[26]研究了传感器网络里数据聚合的能量-延迟平衡问题。假设聚合树的结构为一个已知的先验知识,文献[26]中提出了离线和在线的报文调度机制。报文调度机制是一个迭代的数值优化算法,优化聚合树中传感器的总能耗,但不保证迭代算法的收敛速度。基于动态规划开发出一个伪多项式时间算法。与数据传输完毕就关闭射频天线的经典方法相比,该方法能够获得20%到60%的能量节约。算法的主要

缺点是在动态规划之前,每一个节点都需要等待来自所有子节点的信息,可能增加附加延迟。

3.4 基于网络流量的数据收集算法比较

根据不同的优化目标,出现了很多数据收集算法。这类问题一般是NP的,计算复杂性大,因此提出了很多中心化或分布式的近似算法。从表2可以看出,最大网络生命期数据收集算法一般从构造最优数据收集调度树入手;而相关数据收集算法则充分考虑感知数据在空间或时间上的相关性,利用编码机制减少网络传输量;QoS感知的数据收集算法则试图在低能耗的同时保证一定的网络服务质量。

4 基于移动性的数据收集算法

4.1 存在移动观测者的数据收集算法

这一类网络中引入一个或多个移动数据观测者动态收集数据。移动数据观测者可能是一个具有高功率收发器的移动

机器人或车辆, 可视为移动的 sink。移动数据观测者从基站开始巡游, 穿越网络, 从附近节点收集感知数据, 最后返回远程数据处理中心更新数据。通过利用移动性增强传感器网络的数据传输性能, 大大减少节点的能耗。

射频示踪的斑马^[27]和鲸鱼^[28]被用作在野外环境中收集感知数据的移动节点。这些安装在动物身上的节点在感知区域随机漫步, 仅当相互靠近时才交换数据。网络节点需要随时保持连通, 随机移动路线也难以预测和控制, 不能保证数据的最小延迟。

城市公交车辆上可以加载无线收发器成为移动基站^[29,30]。移动路线和时间可以预测。然而, 数据交换仍然依赖于固有的运输线路, 局限性很大。文献[31]中提出了一种消息摆渡(MF)的方法来研究稀疏 MANET 中的高效数据传输问题, 使用称为消息渡船(message ferries)的移动观测者为传感器部署区域的节点提供数据通信服务。在文献[32-34]中, 利用网格上的二维随机步行模型为称为 MULE 的移动观测者建模。MULE 等概率地移动到四个相邻的网格内。当 MULE 靠近传感器时直接从传感器收集数据, 缓存并传输到有线接入点。这些方法的主要缺点是增加了数据收集延迟, 因为移动观测者不得不走到每一个传感器的通信距离内, 依次收集数据。该机制的性能和代价依赖于数据 MULEs 的数量和传感器的分布。

文献[35]提出了一种数据收集机制, 通过联合考虑移动规划和路由, 最小化传感器的最大平均负载。基于传感器分布服从 Poisson 过程的假设, 传感器的平均负载可以估计为节点密度的一个函数。然而, 当传感器的分布不服从 Poisson 分布时, 平均负载的估计是不准确的。

为了使数据收集机制适合各种网络拓扑, 通过引入称为 SenCar 的移动观测者, 文献[36]中提出一种启发式算法, 能够根据传感器的分布情况动态规划 SenCar 的迁移路径。通过采用负载均衡算法找到转折点, 使网络动态成簇, 可以使 SenCar 避开障碍物, 平衡从传感器到 SenCar 的流量负载, 从而大大延长网络生命期。

目前的研究主要考虑传感器节点静止、只有观测者移动的情况。

4.2 基于移动 Agent 的数据收集算法

移动 Agent 的作用非常类似于移动观测者, 只不过移动的是 Agent 程序而不是节点。移动 Agent 具有传输数据少、能耗低、可伸缩性好、可靠性高等优点, 适合于传感器网络数据收集应用。

在文献[37]中提出了一种相对于 client/server 计算模型的移动 Agent 模型, 并研究了基于该模型的协同信号和信息处理(CSIP)技术。在传感器网络里应用时, 由中心处理节点发送和回收移动 Agents, 携带服务程序的移动 Agent 将在节点之间迁移, 进行数据收集或融合。然后携带部分处理结果继续迁移到下一个节点。该方法可以在提供渐进准确性的同时保持一定程度的容错性, 但是难以估计数据延迟。

文献[38]中使用执行时间、能耗、能量 * 延迟三个指标, 比较了 client/server 模型和移动 Agent 模型的性能。实验结果说明, 当传感器节点数较多时, 移动 Agent 模型的性能较好; 当节点数较少时, client/server 模型的效果好。因此, 提出了基于 Agent 模型和簇结构的混合计算模型, 可以满足 CSIP 的能效、扩展性、容错性的要求。但同样存在延迟难以估计的问题。

通过利用移动观测者或移动 Agent, 可以在大规模网络中实现按需收集数据, 减少卷入数据传输的节点和报文, 同时简化普通节点的设计, 从而大大减少传感器网络的开销。

结束语 由于 WSN 资源有限且应用相关, 尽管如前文所述, 目前国内外已经提出了很多出色的研究方法, 但是该领域还是存在许多有趣的问题与挑战。通过对当前各种数据收集协议和算法的综合分析与总结, 可以看出将来的某些研究策略与发展趋势:

(1) 根据实际的应用要求设计网络的数据传输结构。数据收集协议的性能与网络结构密切相关。不同的应用在节点规模、异构性、可靠性、延迟、能耗等目标上可能存在巨大的差异, 必须根据实际需求构造最佳的网络结构和数据收集算法。

(2) 采用相关编码机制, 尽量减少不必要的冗余通信量。目前研究较多的数据聚合能够减少数据量, 但是精度有限。而感知数据通常存在较强的时空相关性, 采用编码方式压缩感知数据, 或通过空间、时间上的采样, 在特定精度和可靠性保障下抑制不必要的数据传输, 就可以获得良好的节能效果。

(3) 满足特定 QoS 要求。随着 WSN 技术的进步, 越来越多的应用将不再仅仅满足于低能耗, 对低延迟、准确性、可靠性、吞吐量等也有越来越多的要求。这些要求本身可能存在一定的制约关系, 设计数据收集算法时需要充分考虑这些可能相互制约的要求。

(4) 可满足实时在线应用的要求。目前提出的许多优化算法计算量太大, 无法适应 WSN 的实际要求, 将来需要开发更多快速、高效的分布式近似算法。

(5) 满足网络移动性要求。利用移动观测者的数据收集机制, 在数据精度、能耗等方面具有独特的优势。在某些应用中也必须考虑移动性的问题。

(6) 集成安全性的数据收集算法。WSN 本身很容易受到安全威胁, 已有的数据收集协议很少考虑安全问题, 因此需要在设计协议时集成必要的安全机制, 构造安全的数据收集协议。

参考文献

- [1] Duarte-Melo E J, Liu M. Data-gathering wireless sensor networks: organization and capacity. Computer networks: The International Journal of Computer and Telecommunications Networking, 2003, 43(4)
- [2] Intanagonwatt C, Govindan R, Estrin D. Directed Diffusion: A Scalable and robust communication paradigm for sensor networks // Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCOM '00). August 2000
- [3] Krishnamachari B, Heidemann J. Application specific modeling of information routing in wireless sensor networks // Proc. IEEE international performance, computing and communications conference. 2004, 23, 717-722
- [4] Kulik J, Heinzelman W R, Balakrishnan H. Negotiation-based protocols for disseminating information in wireless sensor networks. Wireless Networks, March 2002, 8, 169-185
- [5] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient communication protocol for wireless microsensor networks // Proc. of the 33rd Annual Hawaii Int'l Conf. on System Sciences, Maui: IEEE Computer Society, 2000: 3005-3014
- [6] Younis O, Fahmy S. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on Mobile Computing, 2004, 3(4): 366-379
- [7] Manjeshwar A, Agrawal D P. TEEN: A protocol for enhanced efficiency in wireless sensor networks // Int'l Proc. of the 15th Parallel and Distributed Processing Symp. San Francisco: IEEE Computer Society, 2001: 2009-2015
- [8] Chatterjee S, Havinga P. A Dynamic data aggregation scheme for wireless sensor networks // Proc. Program for Research on

- Integrated Systems and Circuits. Veldhoven, The Netherlands, Nov. 2003
- [9] Lindsey S, Raghavendra C, Sivalingam K M. Data gathering algorithms in sensor networks using energy metrics. *IEEE Trans. Parallel and Distributed Systems*, 2002, 13(9): 924-935
 - [10] Ding M, Cheng X, Xue G. Aggregation tree construction in sensor networks // 2003 IEEE 58th Vehicular Technology Conference, 2003, 4(4): 2168-2172
 - [11] Vaidhyanathan K, Sur S, Naravula S, et al. Data aggregation techniques in sensor networks. Technical Report. OSU-CISRC-11/04-TR60, Ohio State University, 2004
 - [12] Fan K W, Liu S, Sinha P. On the potential of structure-free data aggregation in sensor networks // INFOCOM '06, Barcelona, Spain, 2006
 - [13] Fan K W, Liu S, Sinha P. Structure-free Data Aggregation in Sensor Networks. *IEEE TRANSACTIONS ON MOBILE COMPUTING*, 2007
 - [14] Kalpakis K, Dasgupta K, Namjoshi P. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Computer Networks*, 2003, 42(6): 697-716
 - [15] Xue Y, Cui Y, Nahrstedt K. Maximizing lifetime for data aggregation in wireless sensor networks. *ACM/Kluwer Mobile Networks and Applications (MONET) Special Issue on Energy Constraints and Lifetime Performance in Wireless Sensor Networks*, Dec. 2005: 853-864
 - [16] Hong Bo, Prasanna V K. Maximum lifetime data sensing and extraction in energy constrained networked sensor systems. *Journal of Parallel and Distributed Computing*, 2006, 66(4): 566-577
 - [17] Cristescu R, Beferull-Lozano B, Vetterli M. On Network Correlated Data Gathering // Proc. of IEEE INFOCOM, 2004
 - [18] Cristescu R, Beferull-Lozano B, Vetterli M, et al. Network correlated data gathering with explicit communication: NP-completeness and algorithms. *IEEE/ACM TRANSACTIONS ON NETWORKING*, FEBRUARY 2006, 14(1): 41-54
 - [19] von Rickenbach P, Wattenhofer R. Gathering Correlated Data in Sensor Networks // DIALM-POMC'04, Philadelphia, Pennsylvania, USA, October 2004: 60-66
 - [20] Gupta H, Navda V, Das S R, et al. Efficient Gathering of Correlated Data in Sensor Networks // MobiHoc'05, Urbana-Champaign, Illinois, USA, May 2005: 402-413
 - [21] Deligiannakis A, Kotidis Y, Roussopoulos N. Compressing Historical Information in Sensor Networks // SIGMOD 2004, Paris, France: 527-538
 - [22] Ordonez F, Krishnamachari B. Optimal information extraction in energy-limited wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 2004, 22(6): 1121-1129
 - [23] Upadhyayula S, Gupta S K S. Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced convergecast (DAC) in wireless sensor networks. *Ad Hoc Networks*, 2007(5): 626-648
 - [24] Solis I, Obraczka K. In-Network Aggregation Trade-offs for Data Collection in Wireless Sensor Networks. University of California Santa Cruz. Technical Report, 2004
 - [25] Boulis A, Ganeriwal S, Srivastava M B. Aggregation in sensor networks: An energy-accuracy tradeoff. *Ad Hoc Networks*, 2003 (1): 317-331
 - [26] Yu Y, Krishnamachari B, Prasanna V K. Energy-latency tradeoffs for data gathering in wireless sensor networks. *IEEE INFOCOM*, March 2004
 - [27] Juang P, Oki H, Wang Y, et al. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebra-net. *Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2002
 - [28] Small T, Haas Z. The shared wireless infostation model - a new ad hoc networking paradigm (or where there is a whale, there is a way). *ACM MobiHoc* 2003
 - [29] Chakrabarty A, Sabharwal A, Aazhang B. Using predictable observer mobility for power efficient design of a sensor network // Second International Workshop on Information Processing in Sensor Networks (IPSN). April 2003
 - [30] Pentland A, Fletcher R, Hasson A. Daknet: rethinking connectivity in developing nations. *IEEE Computer*, 2004, 37(1): 78-83
 - [31] Zhao W, Ammar M, Zegura E. A message ferrying approach for data delivery in sparse mobile ad hoc networks. *ACM MobiHoc '04*, 2004
 - [32] Jea D, Somasundara A A, Srivastava M B. Multiple controlled mobile elements (data mules) for data collection in sensor networks // 2005 IEEE/ACM International Conference on Distributed Computing in Sensor Systems (DCOSS'05). June 2005
 - [33] Shah R C, Roy S, Jain S, et al. Data MULEs: modeling a three-tier architecture for sparse sensor networks // IEEE Workshop on Sensor Network Protocols and Applications (SNPA). 2003
 - [34] Jain S, Shah R C, Brunette W, et al. Exploiting mobility for energy efficient data collection in wireless sensor networks // ACM/Kluwer MONET 2005
 - [35] Luo J, Hubaux J-P. Joint Mobility and Routing for Lifetime Elongation in Wireless Sensor Networks. *IEEE INFOCOM* 2005. 2005
 - [36] Ma M, Yang Y. SenCar: An energy efficient data gathering mechanism for large scale multi-hop sensor networks // 2006 International Conference on Distributed Computing in Sensor Systems (DCOSS'06). San Francisco, June 2006
 - [37] Qi Hairong, Xu Yingyue, Wang Xiaoling. Mobile-Agent-Based Collaborative Signal and Information Processing in Sensor Networks // Proceedings of the IEEE, 2003, 91(8): 1172-1183
 - [38] Xu Yingyue, Qi Hairong. Distributed computing paradigms for collaborative signal and information processing in sensor networks. *J. Parallel Distrib. Comput*, 2004(64): 945-959
 - [39] 任丰原, 黄海宁, 林闯. 无线传感器网络. *软件学报*, 2003, 14(2): 1148-1157

(上接第 28 页)

流媒体资源 P2P 节点的策略的优化等。

参 考 文 献

- [1] Naicken S, Basu A, Livingston B, et al. A Survey of Peer-to-Peer Network Simulators // Proceedings of The Seventh Annual Postgraduate Symposium, Liverpool, UK, 2006
- [2] Aberer K, Despotovic Z. Managing trust in a peer-to-peer information system [A] // Proceedings of the 10th International Conference on Information and Knowledge Management (ACM CIKM). New York, USA, 2001
- [3] Freedman M, Vingralek R. Efficient P2P Lookup for a Distributed Trie [A] // Proc. 1st International Workshop on Peer-to-Peer Systems (IPTPS '02) [M]. Cambridge, MA, March 2002
- [4] Schlosser M T, Condie T E, Kamvar S D. Simulating a File-Sharing P2P Network [A] // the Proceedings of 1st Workshop on Semantics in Peer-to-Peer and Grid Computing at the Twelfth International World Wide Web Conference, Budapest, Hungary, May 2003
- [5] Li J, Stribling J, Morris R, et al. A performance vs. cost framework for evaluating DHT design tradeoffs under churn // INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, vol. 1, Mar. 2005: 225-236
- [6] Garca P, Pairet C, Mondjar R, et al. PlanetSim: A new overlay network simulation framework. *Software Engineering and Middleware*, 2005(3437): 123-136
- [7] Saroiu S, Gummadi P K, Gribble S D. A measurement study of peer-to-peer file sharing systems [A] // Proceedings of Multimedia Computing and Networking 2002 (MMCN'02). San Jose, CA, USA, January 2002
- [8] Crespo A, Garcia-Molina H. Semantic Overlay Networks [J]. In Submitted for publication, October 2002
- [9] Adar E, Huberman B A. Free Riding on Gnutella [J]. 2000, 5 (10)
- [10] Ripeanu M, Foster I. Mapping the Gnutella Network - Macroscopic Properties of Large-scale P2P Networks [J]. *IEEE Internet Computing Journal*, 2002, 6(1)
- [11] Stoica I, Morris R, Liben-Nowell D, et al. Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications [J]. *IEEE/ACM Transactions on Networking*, 2003, 11(1): 17-32
- [12] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems // Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware). Nov. 2001
- [13] Zhao B Y, Huang Ling, Stribling J, et al. Tapestry: A Resilient Global-scale Overlay for Service Deployment [J] *IEEE Journal on Selected Areas in Communications*, 2004, 22(1): 41-53