

文章编号: 2095-2163(2023)01-0001-05

中图分类号: TP391.41

文献标志码: A

无线传感器网络中数据收集算法研究综述

陈昆仪, 高宏

(哈尔滨工业大学 计算学部, 哈尔滨 150001)

摘要: 感知数据的采集与传输是无线传感器网络最基本的任务之一, 数据收集算法一直是学术界和工业界关注的重点。近年来, 研究者们针对不同拓扑结构无线传感器网络中不同类型数据收集问题展开了相关研究。本文将针对现有的数据收集算法进行简单介绍, 并对各算法的优缺点进行分析与说明。

关键词: 无线传感网; 数据收集; 算法分析

A review on data collection algorithms for WSN

CHEN Kunyi, GAO Hong

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 Data Collection is one of the fundamental tasks in WSNs and efficient data collection algorithms are always the focus of the academic community. In the past years, researchers have investigated the data collection problem under different network topologies, with different constraints and optimization objectives. In this paper, we review some typical works on the data collection problem and analyze the strengths and drawbacks of these works.

【Key words】 wireless sensor networks; data collection; algorithm analysis

0 引言

随着 WiFi、蓝牙、LoRa、NB-IoT 等无线通信技术的发展以及嵌入式系统、分布式计算系统的成熟, 万物互联、智慧城市的愿景成为可能, 使得物联网 (Internet-of-things) 成为国内外关注的热点。无线传感器网络是物联网的重要组成部分, 一个无线传感器网络往往由几个到几千个无线传感器节点和一个或多个 sink 节点构成, 这些无线传感器节点能够从物理世界中采集温度、湿度、照片等数据, 并将这些数据通过无线传输以一跳或多跳的方式上传到汇聚节点上。无线传感器网络在环境监测、军事监测、交通监测以及结构健康监测等领域具有广泛的发展和前景。

感知数据的采集与传输是无线传感器网络的基本任务, 高效的数据收集一直是学术界与工业界关注的重点问题。受成本、电量和信道资源的限制, 无线传感器节点一般采用 WiFi、蓝牙等短距离通信技

术, 只能与邻近的节点进行通信, 一个感知数据包往往需要经过多次转发才能到达汇聚节点。无线传感器网络中的数据收集, 是指为每个节点安排合适的传输路径以及传输时刻, 保证数据包被 sink 节点成功接收。

在无线传感器网络中, 设计高效数据收集算法时需要考虑的重点在于:

(1) 避免无线传输冲突, 避免重传。在无线传感器网络中收集数据时, 为了提高数据传输效率, 往往令网络中的多个节点同时进行数据传输, 无线信号之间不可避免地会相互叠加和干扰。当信号的信噪比小于一定的阈值时, 接收节点就无法解析信号, 造成丢包和重传。因此, 在设计数据传输算法时, 关键在于保证同时进行的数据传输间不会相互冲突;

(2) 节约能量, 延长网络寿命。在传统有源网络中, 由于电池电量有限, 数据收集算法中的一个重要目标就是最大化网络的寿命;

(3) 利用节点自身的计算功能, 降低需要传输

基金项目: 国家自然科学基金重点项目 (U19A2059); 重点研发 (2019YFB210190)。

作者简介: 陈昆仪 (1994-), 女, 博士研究生, 主要研究方向: 无线传感器网络; 高宏 (1966-), 女, 博士, 教授, 博士生导师, 主要研究方向: 传感器网络、海量数据计算。

通讯作者: 高宏 Email: honggao@hit.edu.cn

收稿日期: 2022-04-18

的数据量。由于传感器节点的传输能耗远大于其计算能耗,一个传感器节点传输 1 比特信息 100 m 距离所需要的能量大约相当于执行 3 000 条计算指令消耗的能量。因此,当计算任务已知时,可以采取“边传输边计算”的方式来降低数据量,进而提高传输效率。

1 传统数据收集算法

在传统有源网络中,已经有一些数据收集算法被提出。如:文献[1]证明了在协议冲突模型下,以最小化延时为目标的数据收集问题是 NP-难的。为解决该问题,文献[2]中对线型网络、树形网络以及一般网络的拓扑分布,给出了集中式的调度算法。文献[3]以同时最小化数据传输的时间利用效率和能量利用效率为目标,在协议冲突模型下,给出了一个分布式数据收集算法。文献[4]分别对树形网络和一般网络,提出了分布式的数据收集算法,并证明:

(1) 对于一个树形网络,至少需要 $\max\{3n_k - 1, N\}$ 个时间槽才能完成一次数据收集(n_k 是以 sink 的一跳邻居为根的最大子树中节点数量, N 是网络中节点数量);

(2) 对于一般的网络拓扑,至少需要 $3N$ 个时间槽才能完成一个数据收集。文献[5]中则提出了第一个在物理冲突模型下的数据收集算法。

为了降低需要传输的数据量,进而提高效率,研究人员提出在节点内使用数据压缩技术来解决此问题。数据压缩技术可以分为无损压缩的数据收集算法^[6]和有损压缩的数据收集算法^[7]。现有的基于无损压缩的数据收集算法,主要依赖压缩感知技术。在压缩感知理论下,当数据在某组基底存在稀疏表达时,数据可以被有效地压缩,并通过解最优化方程的方法,可将原始数据进行精确地恢复^[8]。感知数据往往具有时间相关性和空间相关性,已经有大量的理论和实验表明,感知数据在离散余弦变换、傅里叶变换及小波变换的基底存在稀疏表达^[9]。文献[6]是第一篇将压缩感知技术应用到多跳有源网络中数据收集问题的论文,文中给出的理论分析及实验结果表明,在保证 sink 端能够对原始感知数据进行准确恢复的前提下,可以大大降低网络整体数据传输量,且不会在节点端引入大量的计算代价或复杂的控制信息交换。此外,基于压缩感知理论的数据收集算法可以使网络中各节点的工作负载均衡,有效地延长了网络寿命。为了进一步降低网络

中需要传输的数据量,文献[10]提出了基于混合压缩感知技术的数据收集算法 Hybrid-CS,该算法中只在部分节点执行压缩感知算法。即对于其中的每个节点 v_i ,只有当其转发的原始数据项量大于 $M - 1$ 时(M 为压缩感知因子,取决与感知数据的稀疏性。),或是需要转发的数据中包含压缩感知向量时,才会执行压缩感知算法;否则只转发数据,不做任何压缩处理。而在文献[10]中,只是将 Hybrid CS 的调度问题形式化成一个优化问题,并提出了一个集中式的离线算法来解决这个问题。但由于离线的、集中式的数据收集算法鲁棒性较差,并且需要汇聚节点不断地与各个节点交换信息,因此不适合在现实无线传感网中使用。文献[11]中也提出了一个类似集中式的数据收集算法。该算法将压缩感知技术与数据传输相结合,其目标是最小化总体的能量消耗。并且以两个节点之间的能量消耗作为网络中边的权重,建立一棵以汇聚节点为根的数据收集树,每个节点沿着树上的路径来完成数据传输。

文献[7]提出了基于有损压缩的数据收集算法,其基本想法是将感知数据序列看成时间序列,接着对这些时间序列做压缩。例如:采用数值逼近理论,也就是用多项式函数来拟合感知数据,之后向 sink 节点传输拟合多项式的参数而不是原始的感知数据;或者采用统计学的理论,用 Piece-wise 常数近似方法或主成分分析方式来压缩这些感知数据。

在实际的物联网应用中,收集感知数据的目的是为了下一步的分析和决策,往往需要对感知数据提取特征。但在现有的近似数据收集算法中,只保证了原始传感器数据和解压后的数据的差距小于一个给定的阈值,并没有保证基于原始感知数据提取的数据特征与从解压缩数据提取出来的数据特征之间的差距,很有可能出现由于前面的有损压缩给后续分析带来极大的困难,甚至很可能由于关键信息被抹去,使得后续的分析无法进行。例如,在基于加速度数据的电梯异常检测应用中,相比于正常运行的电梯,故障电梯的加速度会呈现出一些细小的波动。但如果采用近似数据收集的算法,这些细小的波动将会被移除,则无法从近似的感知数据中检测到电梯故障,使之造成事故的发生。

2 以最小化信息年龄为目标的数据收集算法

文献[12]中提出的信息年龄(Age of Information, AoI),是描述信息新鲜性度的量。在许多监测应用中,节点需要向感兴趣的接收方(一般是汇聚节点)发送关

于被监测对象当前状态的数据, 汇聚节点每收到一个监测对象的数据更新包, 就对该对象当前的状态做一次更新。对于一个监测对象 i , AoI 指的是当前时刻 t 与上一次汇聚节点接收到该监测对象感知数据采集时刻 $g^{last}(t)$ 的差值, 即 $AoI_i(t) = t - g^{last}(t)$ 。

近年来, 以最小化 AoI 为目标的数据收集算法, 在传统有源网络中已有许多相关研究结果被发表。在该领域, 从环境中采集关于监测对象的数据, 并生成数据更新包的节点被称为源节点。大多数已有研究针对的均是单跳网络, 即网络中的每个节点都可通过一跳传输, 向汇聚节点发送数据。在此场景下, 其研究的问题可以简化成为每个源节点安排产生数据更新包的时刻, 并且决定何时激活源节点和汇聚节点之间的边, 以进行数据传输。文献[13]中证明: 即使在单跳网络中、在物理冲突模型下, 以最小化 AoI 为目标的数据收集问题是 NP-Hard 的。文献[14]中, 针对最小化峰值 AoI 和最小化均值 AoI 数据收集问题, 提出了一个基于静态调度原则的算法, 并证明了当假定解空间为所有基于静态调度原则算法的前提下, 该调度算法可以实现峰值 AoI 最小化; 对于均值 AoI , 该算法取得的均值 AoI 和最优的 AoI 的差距在一个常数因子之内。在单跳网络中, 除了数据收集问题, AoI 最优化问题也在其它方向被研究, 其中包括: 数据包生成控制^[12]、队列控制^[15] 以及链路调度规则^[14] 等研究。

由于这些研究不需要考虑转发问题, 因此都不能直接被应用到多跳网络中。而转发正是多跳网络中关键所在, 也是其与单跳网络的最大区别。在许多应用中(如: 森林监测系统、地下管道监测系统等), 大量的传感器被分散地部署在广阔空间中, 使得很多节点与汇聚节点之间的距离很远, 在这样的场景下, 每个节点和汇聚节点之间都使用单跳传输的方式通信是不现实的, 这会带来极大的误码率和丢包率。

因此, 研究者们开始在多跳的无源网络中设计以最小化 AoI 为目标的数据收集算法。文献[16]中关注了一个简单的问题空间。假设网络中的每条链路只能遵循某一个固定的概率分布函数被激活, 该如何选择最优的概率分布函数。在此设置下, 文中推导出了最优的概率分布。文献[17]中考虑的情景是: 网络中的节点被分割成固定的对, 每一对中有一个发送者和一个接受者, 发送者周期性地向接收者发送数据。文中研究是在吞吐量下界给定的约束下, 如何最小化峰值 AoI 和均值 AoI 的问题, 并给出了一个集中式的伪多项式时间复杂度算法。文献[18]中考虑的场景是: 网络中的每个节点既是源节点又是监测者, 即每个节点都

在本地维护网内所有节点的 AoI 曲线。在此场景下, 所有节点都会接收到来自其它节点的数据更新包。以此设定, 作者推导出在协议冲突模型以及物理冲突模型下, 峰值 AoI 和均值 AoI 的下界。

然而, 上面提到的研究中都存在明显的缺陷。如: 文献[16]中, 网络中的每一条链路都会以一定的概率被激活, 这个概率服从一个固定的概率分布函数。一个数据包在链路上被成功地传递, 当且仅当在该链路被激活的时刻, 所有与其产生冲突的链路都没有被激活。因此, 在节点密度较大的情况下, 这个调度算法的效率会非常低下。而且, 算法中每个节点到 sink 节点的传输路径都是固定的。当一个节点被损坏后, 这些更新包都会丢失。文献[17]则忽略了对数据包生成时刻的控制, 而数据包的生成时刻直接影响着 AoI 的大小。文献[18]考虑的是一个非常特殊的情形, 而在大多数场景下, 都认为网络中的节点是源节点, 而汇聚节点是唯一的监督者。

3 数据聚集算法

在一些无线传感器网络应用中, sink 所承担的计算任务是已知的。在一些监测系统中, 用户只关注感知数据的统计数据, 如被监测区域的平均温度、加速度的最大值等等。

此外, 由于存在原始感知数据的总量远大于统计数据的量(如 1 000 个温度值的最大值只有一个), 且传感器节点的传输功耗远高于其计算功耗(传输 1 比特信息 100 m 距离所需要的能量大约相当于执行 3 000 条计算指令消耗的能量^[19])的情况, 研究者们提出了将数据收集与计算相融合的方法。即当节点收到来自其它节点的感知数据时, 在节点内部进行聚集操作(如求平均值、最大值等), 以此降低需要传输的数据量, 进而降低传输延时与能耗。上述的处理过程, 称为数据聚集。在传统的有源无线传感器网络中, 最小化聚集延时调度 (Minimum Latency Aggregation Scheduling, MLAS) 是一个经典的问题, 已经有大量的算法被提出。在有源网络中, 默认每个节点在任意时刻都有足够的能量可用于数据传输。在此假设下, MLAS 问题已被证明是一个 NP-难的问题^[20]。文献[20]中构造了一棵以 sink 节点为根的最短路树作为数据聚集树, 并给出了一个延时上界 $\Delta - 1$ 的集中式算法。文献[21]中提出了一种基于平衡的最短路树的算法, 并证明该算法取得了更好的性能。文献[22]提出了一种基于最小联通支配集的数据聚集树, 做调度时采用最先适配 (First-Fit) 原则。在这个设置下, 给出了一个延时上界为

$23R + \Delta - 18$ 的集中式算法,其中 R 表示网络的宽度,即距离基站最远的节点和 sink 节点之间的跳数。文献[22]提出了一个优化的最先适配原则,将时延的上界降到了 $16R + \Delta - 11$ 。文献[23]中提出了一个更先进的基于联通支配聚集树的调度算法,该算法延时的上界为 $1 + O\left(\frac{\log R}{\sqrt{R}}R\right) + \delta$ 。

上述所有算法都是集中式的。然而,集中式的算法很难在现实无线传感器网络中应用。这是因为在实际应用中,传感器节点很容易出现故障,传感器间的连接也非常不稳定,这使得网络的拓扑结构动态变化,如果采用固定的聚集树,则会有两种可能:

(1)假设所有数据聚集周期都使用同样的聚集树,由于网络拓扑变化,树上的链路发生断裂,数据无法被成功地传递到 sink 节点;

(2)如果在每一轮数据聚集前都将节点的拓扑信息传递到 sink 节点,sink 节点执行数据收集算法,再将所有的传输调度方案传递到对应的各个节点。虽然这个方案是可行的,但效率非常低下。因为在每一轮数据聚集前,sink 节点都需要将调度信息发送给每一个节点,这些调度信息的传输会带来大量附加的能量、时间和信道资源的开销。文献[24]提出了第一个分布式算法,并证明了算法聚集延时的上界为 $48R + 6\Delta + 16$ 。随后,文献[25]将此算法做了进一步改进。

在占空比传感网中,每个节点按照一定的周期休眠或工作,MLAS 问题已在许多文献中被研究。如:文献[26]研究了在每个调度周期里,每个节点只工作在一个时间槽的情况,并提出一个集中式的算法。文献[27]中,假设每个节点在一个调度周期内有多个时间槽可用于工作,并提出一个时延上界为 $(15R + \Delta - 3) \times T$ 的算法 (T 是一个聚集周期的时间长度)。文献[28]提出了一个分布式算法,在此一个节点的所有邻居都是其候选父节点。实验结果表明,此算法优于前两种算法。然而,由于无源网络中的延时主要来自于无源节点能量的不足,因此这些算法都不能直接用于无源传输网络。

4 结束语

研究者在无线传感器网络中的数据收集调度问题上已深耕多年,针对不同的优化目标提出了多种有效的算法。对于这些算法进行了总结,对其优缺点进行了分析。总而言之,降低时延,减少能耗,提高网络吞吐量依然是无线传感器网络中数据收集算法的核心优化目标。

参考文献

- [1] CHOI H, WANG J, HUGHES E A. Scheduling for information gathering on sensor network[J]. *Wireless Networks*, 2009, 15(1): 127-140.
- [2] FLORENS C, FRANCESCHETTI M, MCELIECE R J. Lower bounds on data collection time in sensory networks[J]. *IEEE Journal on selected Areas in Communications*, 2004, 22(6): 1110-1120.
- [3] SONG W Z, YUAN F, LAHUSEN R. Time-optimum packet scheduling for many-to-one routing in wireless sensor networks[C]//2006 IEEE International Conference on Mobile Ad Hoc and Sensor Systems. IEEE, 2006: 81-90.
- [4] GANDHAM S, ZHANG Y, HUANG Q. Distributed time-optimal scheduling for convergecast in wireless sensor networks[J]. *Computer Networks*, 2008, 52(3): 610-629.
- [5] GONG D, YANG Y. Low-latency SINR-based data gathering in wireless sensor networks[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(6): 3207-3221.
- [6] LUO C, WU F, SUN J, et al. Compressive data gathering for large-scale wireless sensor networks[C]//Proceedings of the 15th annual international conference on Mobile computing and networking. 2009: 145-156.
- [7] CHENG S, CAI Z, LI J. Approximate sensory data collection: A survey[J]. *Sensors*, 2017, 17(3): 564.
- [8] CANDÈS E J, WAKIN M B. An introduction to compressive sampling[J]. *IEEE signal processing magazine*, 2008, 25(2): 21-30.
- [9] QIN Z, FAN J, LIU Y, et al. Sparse representation for wireless communications: A compressive sensing approach[J]. *IEEE Signal Processing Magazine*, 2018, 35(3): 40-58.
- [10] LUO J, XIANG L, ROSENBERG C. Does compressed sensing improve the throughput of wireless sensor Networks? [C]//2010 IEEE International Conference on Communications. IEEE, 2010: 1-6.
- [11] XIANG L, LUO J, ROSENBERG C. Compressed data aggregation: Energy-efficient and high-fidelity data collection[J]. *IEEE/ACM transactions on Networking*, 2012, 21(6): 1722-1735.
- [12] KAUL S, YATES R, GRUTESER M. Real-time status: How often should one update? [C]//2012 Proceedings IEEE INFOCOM. IEEE, 2012: 2731-2735.
- [13] HE Q, YUAN D, EPHREIMIDES A. Optimal link scheduling for age minimization in wireless systems [J]. *IEEE Transactions on Information Theory*, 2017, 64(7): 5381-5394.
- [14] TALAK R, KARAMAN S, MODIANO E. Optimizing information freshness in wireless networks under general interference constraints [J]. *IEEE/ACM Transactions on Networking*, 2019, 28(1): 15-28.
- [15] COSTA M, CODREANU M, EPHREIMIDES A. Age of information with packet management[C]//2014 IEEE International Symposium on Information Theory. IEEE, 2014: 1583-1587.
- [16] TALAK R, KARAMAN S, MODIANO E. Minimizing age-of-information in multi-hop wireless networks[C]//2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2017: 486-493.
- [17] LIU Q, ZENG H, CHEN M. Minimizing age-of-information with throughput requirements in multi-path network communication[C]//Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing. 2019: 41-50.

(下转第11页)