Vol. 18 No. 1 Jan. 2008

无线传感网最大剩余能量索引数据分发算法

陶孜谨,龚正虎,欧阳一星,徐金义 (国防科学技术大学 计算机学院,湖南 长沙 410073)

摘 要:提出一种基于索引和局部存储的(Index and Local Storage - based, ILS)数据分发算法 MREIB - DD。对于 ILS 类型的数据分发算法,一个事件的监测数据被存储在该数据的监测节点或监测节点的邻居节点。一个存储节点仅当接收到一个来自 Sink 的查询,才把监测数据发送至 Sink。MREIB - DD 算法选择网络中有最大剩余能量的节点存储索引信息,传感器节点监测到数据时向索引节点发送该数据的有关索引信息。用户的查询信息先到达索引节点,索引节点把查询转发到数据存储点,存储点对查询进行响应。此算法避免了感知数据的网内传输和查询泛洪带来的开销,分析表明该算法性能优于 GHT - DCS 算法而复杂度增加较少,是能量高效的数据分发算法。

关键词:无线传感器网络;索引存储;数据分发;剩余能量;MREIB;能量高效

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2008)01-0017-05

Maximum Residual Energy Index Based Data Dissemination Algorithm for Wireless Sensor Networks

TAO Zi-jin, GONG Zheng-hu, OUYANG Yi-xing, XU Jin-yi (School of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: A data dissemination algorithm (MREIB – DD) which is based on index and local storage (ILS) is proposed. For ILS, the sensed data are stored in the node which has sensed the data or the neighboring node. The storage node sends the data to the sink only when it has received a query. MREIB – DD selects the node which has the largest residual energy as the index storage node. When a data is sensed, the index of the data is sent to the index node. The user query sent from the sink is sent to the index node and directed to the node which stores the data. The data storage node responds the query with the data. The algorithm avoids the in – network data transferring and the query flooding. The performance analysis of the algorithm is given and compared to the GHT algorithm. The new algorithm can improve the energy – efficiency and the complexity of the algorithm has not been increased much, so the new algorithm is an energy – efficient data dissemination algorithm for the WSNs.

Key words: wireless sensor networks; index storage, data dissemination; residual energy; MREIB; energy - efficient

1 相关工作

随着传感器技术、嵌入式技术以及低功耗无线通信技术的发展,生产具备感应、无线通信以及信息处理能力的微型无线传感器已成为可能。这些廉价的、低功耗的传感器节点共同组织成无线传感器网络^[1~3],它有广阔的应用前景,可以应用在国防军事、救灾、环境监测等各个领域。由于传感器节点的能量是有限的,所以设计能量有效的数据分发算法是目前研究热点之一。

在过去几年内,已有很多关于无线传感器的数据分发算法^[4~9],一种为基于外部存储的数据分发算法 (ES),它依赖于无线传感器网络外部用于收集和存储感知数据的基站(base station),当无线传感器网络中节点发出很多查询时,此算法是低效的^[9],因为数据在传感器节点和基站之间被反复发送。文献[4]提出了以数据为中心存储的数据分发算法(DCS),一旦源节点监测到事件便将相关数据存储至特定节点,这种算法适应性差。

为了避免存储感知数据带来的开销,研究者提出了基于局部存储的数据分发算法,直接扩散(directed diffusion, DD)^[8,10]和 TTDD(two-tier data dissemination)^[9]是基于 LS 的两种典型的数据分发算法。在DD算法中,源节点一旦监测到事件,便直接存储数据至本地存储点。终端节点 Sink 需要数据时产生一定

收稿日期:2007-06-15

基金项目:国家 973 重点基础研究发展规划项目(2003CB314802) 作者简介:陶孜谨(1973-),男,湖南长沙人,助理研究员,博士研究 生,研究方向为计算机网络;龚正虎,教授,博士生导师,研究方向为 计算机网络。

第18卷

的查询分组,并将查询分组在整个网络中洪泛,直至到 达所有存储点,存储点根据查询分组将匹配数据返回 至终端节点。考虑到一个传感器网络中有大量节点, 在网络范围内洪泛,必将带来巨大的通信开销;而对于 TTDD算法,在其运行之初要建立一个坐标网格,这需 要大量的数据通讯,而且网格的维护也需要一定的开销,这与传感器网络拓扑变化大、节点能量有限的特点 相违背。

为解决上述算法的缺点,笔者提出了一种基于索引和局部存储的(Index and Local Storage - based, ILS)数据分发算法 MREIB。对于 ILS 类型的数据分发算法,一个事件的监测数据被存储在该数据的监测节点或监测节点的邻居节点(这些节点叫存储节点)。一个存储节点仅当接收到一个来自 Sink 的查询,才把监测数据发送至该 Sink。索引信息被存储至网络内索引点,用户通过 Sink 节点发送的查询信息先到达索引节点,索引节点根据自身存储的信息把查询转发到数据的存储节点,存储节点对查询进行应答。此算法避免了存储监测信息和查询洪泛带来的开销。

2 基于索引和局部存储的数据分发算法

2.1 系统模型

无线传感器网络内的节点实时监测多个目标,并定期产生感知数据,用户通过 Sink 不时地向网络内节点发出关于某目标事物当前状态或某段时间内活动的查询,这里假设传感器节点是静止的,并且通过 GPS^[3,11,12]或其它技术可获取自已的地理位置。该算法是建立在 Greedy(贪婪算法)^[13]路由算法之上的。

在此算法中,监测目标被分类,每一类有一个属性值关键字k,把监测数据进行本地存储。把监测信息进行本地存储之后,形成包含k和该节点位置信息以元数据形式发送至索引点作为索引信息进行存储,即整个网络覆盖区域内只有一个索引点。

当 Sink 发出关于 k 类型数据查询时,则查询分组 先到达索引点获取 k 类型数据的存储位置,再把查询 分组发送至存储点,存储点返回数据给 Sink。如图 1 所示。

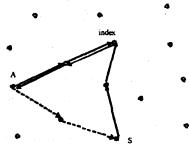


图 1 ILS 数据存储原理

2.2 数据分发算法的比较

假设传感器网络由 n 个节点组成,目的是监测 T 类不同事件。采用 D_{coal} 代表产生的所有监测数据的分组数量,Q 代表查询的命令数量, D_q 代表查询返回的数据分组的数量,并简单假设所有分组的大小相等。

用总通信量和热点通信量[4] 作为性能分析的评估指标。由于存储点或终端节点的邻节点是整个传感器网络中负载最大的点,故对这些节点上通信量的评估可作为热点通信量。按照文献[4] 中参数计算,并假设所有的元数据分组为 $D_{\text{data-m}}$, ILS 算法中数据存储的通信复杂度为0,位置信息发布的通信复杂度、查询分组发送至存储点和存储点返回数据的复杂度都为 $O(\sqrt{n})$ 。

表	1	数据	分	发	算	法	的	比较

算法类型	总通信量	热点通信量
ES	$D_{\text{total}} \sqrt{n}$	D _{total}
LS	$Qn + D_q \sqrt{n}$	$Q+D_q$
DCS	$Q\sqrt{n} + D_{\text{total}}\sqrt{n} + D_{q}\sqrt{n}$	$Q + D_q$
ILS	$D_{\text{dece-an}}\sqrt{n} + Q\sqrt{n} + D_q\sqrt{n}$	

由表 1 可知:如果所有其他参数固定不变,则当网络规模很大即 n 很大时,LS 算法总通信量最大,它与ES 算法总通信量的比值为: $\frac{Q\sqrt{n}}{D_{\text{total}}} + \frac{D_{q}}{D_{\text{total}}}$; DCS 算法的总通信量恒大于 ES 算法,且二者之比为 $1 + \frac{Q+D_{q}}{D_{\text{total}}}$,若监测数据量大,则该比值不大(假设每次查询至少会返回一个数据,则该比值的最大值为 3,最小值近似于 1);若监测数据量远大于查询数据量,即 $D_{\text{total}} \gg \text{Max} \lfloor Q, D_{q} \rfloor$,则 LS 算法和 DCS 算法的热点通信量远小于ES 算法。由于 D_{total} 是恒大于 $D_{\text{data-mn}}$ 的,所以 DCS 与 ILS 的差值: $(D_{\text{total}} - D_{\text{data-mn}})\sqrt{n}$ 也恒大于零。故若仅从总通信量和热点通信量来看:

- 1) 若监测数据量与查询数据量相似,即 $D_{\text{total}} \approx \max \lfloor Q, D_q \rfloor$,则 ES 最优。
- 2) 网络规模中等, 监测数据量远大于查询数据量,同时监测数据量大于系统查询规模, 即 $(D_{total}\gg Max \lfloor Q, D_q \rfloor) \land (D_{total}>Q\sqrt{n})$,则 LS算法最优。
- 3) 网络规模大, 监测数据量远大于查询数据量,同时监测数据量小于系统查询规模,即 $(D_{total} \gg Max \lfloor Q, D_a \rfloor) \land (D_{total} \leq Q \sqrt{n})$,则 DCS 算法最优。
- 4) 网络规模大,监测目标类型较少,而监测数据量较多,且元数据量小于查询规模,即($D_{total}\gg Max \lfloor Q, D_q \rfloor$) \wedge ($D_{data-an} \leq D_q$),则 ILS算法最优。

维普资讯 http://www.cqvip.com

3 基于最大剩余能量索引的数据分发算法

3.1 MREIB-DD 基本原理

MREIB - DD (Maximum Residual Energy Index Based - Data Dissemination,基于最大剩余能量索引的数据分发)算法是基于 ILS 类型的。它沿用了以数据为中心的路由算法的基本思想,即查询分组和监测数据都采用元数据标识,通过分配不同的属性值来表示不同的任务或监测事件。图 2 描述了 MREIB - DD 算法的数据分发原理。

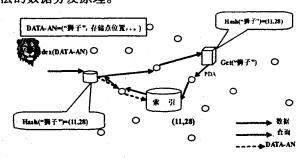


图 2 MREIB-DD 数据分发原理

图中当监测节点 1 监测到属性值关键字为"狮子"的信息后,把监测信息进行本地存储,并把监测信息的属性值关键字 k、位置和剩余能量以元数据形式向索引点发送,索引节点对索引表进行相应的更新,并把收到的元数据再发给索引备份点。 Sink 节点需要查询"狮子"的信息时,由于位置已知,Sink 先向离它最近的邻居节点 A 发出一个查询信息,目的是为了获取当前索引节点的位置,获取当前索引点位置信息后,便向索引点发送属性值关键值为"狮子"的查询分组。索引点收到查询分组后,通过查询索引表,找到所需信息的存储位置为监测点 1,再把查询分组转发至存储点(即监测点 1),存储点找到与查询分组所匹配的监测数据,并通过 Greedy 路由算法发送至 Sink 节点。

3.1.1 数据分组存储

当网络中节点监测到新类型数据时, MREIB-DD 把这个监测信息进行本地存储, 并把这个节点监测到信息的属性值关键字 k、位置和剩余能量, 以元数据形式发给索引节点, 这个元数据在传输到索引节点的过程中, 每经过一个节点, 则把这些转发节点的位置和剩余能量也作为一项信息加入到这个元数据中, 最后当分组到达索引节点时, 把元数据中所记录的全部信息对应到索引列表中相应的节点信息进行更新。

3.1.2 索引点的选择

系统初始时,通过预先定义的 Hash 函数,例如GHT,对检测到的事件类型进行 Hash 来决定初始索引点的位置,并且假定检测区域内的所有节点都知道该映射,这样对于某个特定类型的检测事件就有一个

一致的初始索引点。而不同类型的数据根据 Hash 值的不同可能得到不同的索引点,这样就能避免部分节点通信过热的问题,均衡了系统的负载。

为了避免索引节点因能量耗尽而失效,设一个阈值 S,当索引节点检测到自身的剩余能量小于等于阈值 S时,则开始重新选择索引节点,索引点在自己索引表中检索能量最大的节点,如果有两个或两个以上的节点都具有最大剩余能量时,选择地理位置距离原索引点最近的那个节点作为等待验证的候选索引点。

选出等待验证的新索引点之后,发送一个剩余能量查询信息,询问该节点的剩余能量是否大于另一个阈值 M(M>S)。因为当一个节点监测到新类型数据时,才发送一个分组给索引节点,当该节点监测的数据类型已存在于本节点,则该节点只进行存储,而不向索引节点发送一个报告信息(即:监测到信息的属性值关键字 k、位置和剩余能量),所以在这个过程中,存储会消耗一部分能量,但索引节点中记录的该节点的剩余能量没有及时更新,有可能存在以下问题:索引节点中记录的该节点的剩余能量已小于阈值 M,所以需要一个剩余能量的确认过程。

等待验证的候选索引点接收到此查询信息后,返回自己的剩余能量值(这个返回值向索引节点传递过程中也不断加入了转传节点的剩余能量值,以便在索引点作相应的更新)。若该值小于 M,则在索引节表中重新选择具有最大剩余能量的节点,向其发送剩余能量查询信息,直至找到一个节点的剩余能量大于 M,否则这一查找新索引节点的过程将一直进行下去。如找到满足以上条件的节点,此等待验证的新索引节点被认为是真正的新索引节点,这时,原索引节点把自己的索引表通过 Greedy 路由算法全部发给新索引节点,与此同时,也在网络区域内发送一个广播信息给其它的非索引节点,告知它们新索引节点的位置,当这些节点下次监测到新类型的数据时,把元数据信息发送到新索引节点。

索引点在选择新索引点的过程中,可能需要多次验证新索引点的剩余能量情况,为了避免过多地选择浪费能量,设定一个选择次数的上限 P_s ,因此假设 P_s 为在时间 T 内找到等待验证的新索引节点的次数 P_s $\leq P_s$ \rangle 。

当确定了新索引节点后,原索引点把存储在自身的全部索引信息发送给新索引节点,并且新索引点发一个广播信息给网络内的其它节点,告之自己的位置,当其它节点监测到新类型的数据时把含有存储点位置信息的元数据发给新的索引节点,而不再发给原来的

索引节点。由于广播信息在网络内发送存在一定时延,很有可能原索引点发出广播信息后,依然收到某些节点发来的元数据,所以设定一个时间段t,在原索引点发送广播信息时,开始计时,把t时间段内收到的元数据转发至新索引点。

3.1.3 索引备份点的选择

为了防止索引点突然失效,引入了索引备份点。索引点从其邻居节点中选择剩余能量最大作为索引备份点,把索引表发送至索引备份点,当索引点收到元数据后,则定期将本节点存储的索引信息发送至索引备份点,如果在一段时间内,索引节点每隔时间 a,索引点向索引备份点发送一个 alive 信息,目的是为了告知索引备份点,自己还没有失效。如果在时间 a 内索引点向备份点发送了数据信息,则取消本次 alive 信息的发送,因为此时已经证明索引点仍然活跃,没有必要再发送 alive 信息。

当索引点能量低于 S 时,如 3.1.2 节所述找到新索引点,新索引点找到其索引备份点,经过 t 后,原索引点和原索引备份点清空索引表,释放传感器存储空间。若索引备份点超过时间 a 还没有收到来自索引点的 alive 信息,则认为索引点已失效,此时,索引备份点成为新的索引点,并自如前所述重新选择新的索引备份点。需要注意的是这种情况只发生在索引点突然失效的情况,正常情况下索引点会在能量耗尽之前把数据传递到新索引点,所以一般不会发生此类事件,备份节点是对系统健壮性的一个保证,避免数据因意外情况丢失。

3.1.4 数据分组获取

当 Sink 节点向索引点发出查询分组时,由于原索引点可能因能量低已更换,所以发送查询分组前,应该先确认一下当前索引点的位置。由于节点都有 GPS 定位系统,所以终端节点可以得知离它最近的节点的位置,于是,且又因为索引点更换时,向整个网络区域发送了广播信息告诉所有节点新索引点的位置,所以 Sink 节点便询问离它最近节点,以获取当前索引节点的地理位置,当该最近节点返回索引点位置信息后, Sink 节点便向索引节点发送查询分组,索引节点根据索引列表,把查询分组转发给相应的存储节点,当存储节点收到查询时,把符合查询的存储于本地的监测信息直接发送到 Sink 节点。

3.2 性能分析

3.2.1 MREIB-DD 总通信代价

网络覆盖区域 A 的各参数说明如表 2 所示。在 MREIB - DD 中,网络区域中所有节点各自监测到的 数据类型个数的总数 $D_{deta-an}$,所以有 $D_{data-an}$ 个数据分

组发送至网络区域中索引节点的通信开销为: $(c\sqrt{N})D_{\text{data-an}}d_a$,彼此互为邻节点,新索引节点向其索引备份点发送索引表的开销为 $D_{\text{data-an}}d_a$,这一过程总通信开销为 $(c\sqrt{N}+1)D_{\text{data-an}}d_a$ 。其中 $c\sqrt{N}$ 是源节点发送数据至索引节点路径的平均节点数 $(0< c<\sqrt{2})$ 。

在时间段 T内,索引点每隔时间 a,向索引备份点发送信息的开销为 $\frac{T}{a}d_t$ 。当索引节点找到一个等待验证的候选索引节点,想判定其是否可成为真正的新索引节点时,发出一个剩余能量查询信息,并且这个等待验证的新索引节点对索引节点查询的返回信息的通信开销为 $c\sqrt{N}(d_p+d_r)$,如果在时间 T内,发生了 P次选择,则总的通信开销为 $Pc\sqrt{N}(d_p+d_r)$ 。在时间 T内,真正成为新索引节点的节点数为 V(等待验证的新索引点P-V次没有成为真正的新索引点),原索引节点把已存储在自身的索引信息发送给新索引节点的通信开销为 $c\sqrt{N}d_n$,并在网络内进行广播的通信开销为 Nd_b ,那么这 V 次总的通信开销为 $V(c\sqrt{N}d_n+Nd_b)(d_n$ 是历来所有 D_{durant} 的和)。

表 2 MREIB - DD 参数列表

参数	描述
N	区域 A 中节点数
М	节点分布密度
d	网络内监测信息的大小
d_a	节点向索引节点发送关于新类型数据的元数据的大小
d_m	查询目前索引节点位置的分组大小
d_x	返回给 Sink 目前索引节点位置的信息大小
D _{data-an}	所有节点发现新的属性类型总数
R * R	网络范围
d_p	索引节点向被选节点发送一个剩余能量查询信息的大小
d_q	终端节点从所有节点接收到的数据分组的个数
d_b	原索引节点在网络内广播信息的大小
d _r	被选节点对剩余能量查询的一个返回信息的大小
d_n	存储在索引节点中全部的索引信息的大小
d_l	终端节点共发送查询分组的大小
Q	终端节点共发送查询分组的个数
P	选择新节点的次数
d_t	索引点发送至索引备份点的更新信息
b	成功选择新节点的次数

对于终端节点,向索引节点发送 Q 个查询分组时,也需要进行 Q 次索引点位置的获取,通信开销为: $Q(d_m+d_x+c\sqrt{N}d_l)$,索引节点通过查询索引表找到所需信息的存储位置后,向该节点发送查询元数据,这一过程的开销为 $c\sqrt{N}Qd_l$,所以这一整个查询过程的总通信开销为 $Q(d_m+d_x+2c\sqrt{N}d_l)$ 。

终端节点共获取 d_a 个数据分组的通信开销为:

 $c\sqrt{N}d_{q}d_{\circ}$

MREIB - DD 在时间 T 内总通信量为:

$$\lceil CO_{\text{MREIB-DD}} \rceil \approx (c \sqrt{N} + 1) D_{\text{dtat-an}} d_a + \frac{T}{a} d_t + Pc$$

$$\sqrt{N} (d_p + d_r) + V(c \sqrt{N} d_n + N d_b) + Q(d_m + d_x + 2c \sqrt{N} d_t) + c \sqrt{N} d_t d$$

$$(1)$$

3.2.2 GHT 和 MREIB - DD 最大总通信开销比较GHT 算法的最大总通信开销(CO) 为:

 $V(c\sqrt{N}d_n + Nd_b) - Q(d_m + d_x + c\sqrt{N}d_l)$ (3) 所以 MREIB – DD 的总通信量小于 GHT。

4 总结及下一步工作

以无线传感器网络中数据分发算法的研究与改进 为出发点,研究如何解决无线传感器网络中数据分发 的能源有效性和负载均衡性的问题,在分析了无线传 感器网络工作特点和现有数据分发算法的基础上,提 出了基于 ILS 的最大剩余能量索引数据分发算法。对 其性能进行了理论分析,证明其性能优于已有的 GHT 算法,能降低系统的能量消耗,延长系统的生存时间。 下一步的工作主要是根据无线通信的理论(能量)模型,实际分析这两种算法,在特定应用背景下(例如楼 宇监控)采集各种数据信息,计算其能量消耗的具体 值,与 GHT 算法进行对比。并将此理论应用到实际的 传感器节点上(例如 MicaZ)。

参考文献:

- [1] Akkaya K, Younis M. A Survey on Routing Protocols for Wireless Sensor Networks [J]. in the Elsevier Ad Hoc Network Journal, 2003(8):325-349.
- [2] Akyildiz I, Su W, Sankarasubramaniam Y, et al. Wireless Sen-

- sor Networks: A Survey[J]. Computer Networks, 2002, 38 (4):393 422.
- [3] 孙利民,李建中,陈 渝,等.无线传感器网络[M].北京: 清华大学出版社,2005.
- [4] Ratnasamy S, Karp B, Yin L, et al. GHT: A Geographic Hash Table for Data Centric Storage [C]//ACM International Workshop on Wireless Sensor Networks and Applications. [s. l.]; [s.n.], 2002.
- [5] Ghose A, Grobklags J, Chuang J. Resilient data centric storage in wireless ad hoc sensor networks [C]//Proceedings the 4th International Conference on Mobile Data Management (MDM'03), [s.1.]: [s.n.], 2003; 45 62.
- [6] Greenstein B, Estrin D, Govindan R, et al. DIFS: A Distributed Index for Features in Sensor Networks [C]//First IEEE Ineternational Workshop on Sensor Network Protocols and Applications. [s. l.]: [s. n.], 2003.
- [7] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy Efficient Communication Protocal for Wireless Microsensor network[C]//Proc. of the Hawaii International Conference on System Sciences. [s. l.]:[s. n.],2000.
- [8] Intanagonwiwat C, Govindan R, Estrin D. Directed Diffusion: A Scalable and Robust Communication [C]//IEEE Mobi – COM '00. [s.l.]: [s.n.], 2000.
- [9] Ye F, Luo H, Cheng J, et al. A Two Tier Data Dissemination Model for Large – scale Wireless Sensor Networks[C]//ACM International Conference on Mobile Computing and Networking (MOBICOM'02). [s.l.]; [s.n.], 2002:148 – 159.
- [10] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks[C]//In: Pickholtz R, ed. Proc. of the 6th Annual Int'l Conf. on Mobile Computing and Networking. New York: ACM Press, 2000:56 67.
- [11] US Naval Observatory (USNO). GPS Operations [EB/OL]. 2001 04. http://tycho. usno. navy. mil/gps. html.
- [12] Chong Chee Yee, Kumar S P. Sensor Networks: Evolution, Opportunities, and Challenges [J]. Proceedings of the IEEE, 2003,91(8):1247-1256.
- [13] Finn G G. Routing and addressing problems in large metropolitan scale internetworks [R]. [s. l.]: Information Sciences Institute, 1987.

(上接第 16 页)

- graphCQ: Continuous data flow processing for an uncertain world[C]//In Proc. of the 1st Conf. on Innovative Data Systems Research. Asilomar, CA, USA:[s. n.],2003:269 280.
- [4] Carney D, Centinternel U, Cherniack M, et al. Monitoring streams—a new class of data management applications[C]// In Proc. of the 28th Int'l. Conf. on Very Large Data Bases. Hong Kong: [s. n.], 2002:215—225.
- [5] Arasu A, Babcock B, Babu S, et al. STREAM: The Stanford Data Stream Management System [EB/OL]. 2004. http://

- dbpubs. stanford. edu/pub/2004 20.
- [6] Arasu A, Babu S, Widom J. CQL: A language for continuous queries over streams and realations [C]//In 9th Int'l. Workshop on Database Programming Languages. Potsdam, Germany: [s. n.], 2003:1-11.
- [7] Arasu A, Babu S, Widom J. The CQL Continuous Query Language: Semantic Foundations and Query Execution [R/OL]. 2003 - 10. Stanford University. http://dbpubs. stanford.edu/pub/2003 - 67.