

MR-MC 无线传感器网络基于森林的数据收集研究

张伟平¹, 郭亚红², 王蒙^{1,3}, 倪林雨^{1,3}, 李金宝^{1,3}

(1. 黑龙江大学计算机科学技术学院, 黑龙江 哈尔滨 150080;

2. 黑龙江大学信息科学与技术学院, 黑龙江 哈尔滨 150080;

3. 黑龙江省数据库与并行计算重点实验室, 黑龙江 哈尔滨 150080)

摘要: 传感器网络的部署环境以及节点自身的限制, 导致传感器节点很容易出现故障并且难以维护。在基于树的数据收集过程中, 节点故障或者链路拥塞会造成较高的通信时延, 甚至数据丢失。针对该问题提出以森林作为路由结构进行数据收集的策略。首先提出一个建立森林的算法, 然后以多棵树作为路由结构进行数据收集。理论分析和实验结果表明, 提出的方法可以有效减少数据收集过程中的数据丢失, 在有 25 个故障节点的情况下, 3 棵树的森林路由结构收集的数据量与基于连通支配集的路由树收集的数据量相比多 55%, 并且能降低数据收集的延迟。

关键词: 无线传感器网络; 路由树; 数据收集; 延迟

中图分类号: TP212

文献标识码: A

Forest based data collection in MR-MC wireless sensor networks

ZHANG Wei-ping¹, GUO Ya-hong², WANG Meng^{1,3}, NI Lin-yu^{1,3}, LI Jin-bao^{1,3}

(1. School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China;

2. School of Information Science and Technology, Heilongjiang University, Harbin 150080, China;

3. Key Laboratory of Database and Parallel Computing of Heilongjiang Province, Harbin 150080, China)

Abstract: The limit of node itself and deployment environment of WSN result in the node was prone to failure and difficult to maintain. In the tree-based data collection process, the node failure or link congestion could result in higher communication delay, or even data loss. To solve this problem, a strategy for data collection was proposed which used forest as the routing structure. Firstly, an algorithm for the construction of forest was proposed, and then collect data through trees in the forest. Theoretical analysis and simulation results show that, the method could reduce the loss of data in the data collection process effectively, in the case of 25 fault nodes, the amount of data collected by forest routing structure of 3 trees compared to the amount of data collected from the connected dominating set is more than 55%, and reduce the latency of data collection.

Key words: WSN, routing tree, data collection, latency

1 引言

近年来, 无线传感器网络因其巨大的潜力被广泛应用于军事领域、环境监测、医疗和工农业等领域中^[1]。在这些应用中, 大量的传感器节点监测周

围的环境, 并将感知的数据通过多跳路由传输到汇聚节点。作为无线传感器网络的主要功能之一, 数据收集问题受到众多研究者的广泛关注。

无线传感器网络通常部署在无人值守的野外环境中, 网络中的传感器节点有可能因为出现故障

收稿日期: 2015-10-30; 修回日期: 2016-01-18

通信作者: 郭亚红, jbli@hlju.edu.cn

基金项目: 国家自然科学基金资助项目(No.61370222, No.61300225); 黑龙江省自然科学基金资助项目(No.F201324); 黑龙江省高校科技创新团队建设计划基金资助项目(No.2013TD012); 哈尔滨市优秀学科带头人基金资助项目(No.2015RAXXJ004)

Foundation Items: The National Natural Science Foundation of China (No.61370222, No.61300225), The Natural Science Foundation of Heilongjiang Province (No.F201324), Technology Innovation of Helongjiang Educational Committee (No.2013TD012), The Program for Group of Science Harbin Technological Innovation Found (No.2015RAXXJ004)

而导致无法正常工作,例如节点的电池能量耗尽、动物踩踏以及大雨雷电的破坏等。由于在无线传感器网络中进行数据收集通常使用树作为路由结构,因此,当网络中的某个节点出现故障时,以该节点为根的子树上的所有数据都将无法传输到汇聚节点。

传感器节点使用无线信道进行通信,多条链路竞争通信资源有可能会发生传输失败。链路调度即为每条链路分配指定的传输时槽进行通信,可以提高链路的并行性,有效降低数据收集延迟。目前的链路调度方案均假设在满足特定的干扰模型(协议干扰模型、物理干扰模型、信噪比模型等)下,每次链路调度都是成功的。然而在实际情况中,链路通常需要多次传输才能够成功,造成这种现象的原因有多种,例如链路可能会由比特误码率等原因阻塞。在以树作为路由结构的链路调度中,如果某条链路阻塞,那么该链路不会被调度直到下一个周期的调度时槽到达。此外,如果2个节点使用某个信道进行通信时出现阻塞,那么该信道在最近一段时间内都会处于阻塞状态。因此,如果链路的性能不稳定,那么数据收集的延迟将会受到很大影响。

综合上述2个方面的分析,考虑节点出现故障以及链路阻塞等原因,本文提出了一个新颖的基于协议干扰模型的以森林作为路由结构的数据收集策略。

2 相关工作

对于不同的应用环境,无线传感器网络中数据收集协议的设计目标也不尽相同,下面分别从容量和延迟等方面介绍数据收集协议的研究现状。

文献[2~6]研究了在不同类型的网络中数据收集容量的问题。其中,文献[2]分析了在任意单 radio、单信道无线传感器网络中的数据收集容量问题,该文献分别讨论了当通信模型采用磁盘图模型和一般图模型,干扰模型分别采用协议干扰模型以及物理干扰模型时数据收集容量的上下界。文献[3]和文献[4]分别研究了在 DR-MC 无线传感器网络和大规模概率无线传感器网络中的数据收集容量问题,并分别设计了快照数据收集算法以及连续数据收集算法。文献[5]和文献[6]分别研究了随机网络和异步网络中的数据收集容量问题。

文献[7~12]以缩短数据收集的延迟为目标设计数据收集协议。其中,文献[7]考虑的是建立一个低

延迟的数据收集结构,文献[8]和文献[9]以树作为路由结构,通过调度链路节约数据收集时间。针对收集决策信息问题,当时间不足以收集来自网络中的所有节点的决策时,文献[10]考虑收集具有更高可靠性的决策。文献[11]针对城市建筑中使用的无线传感器网络设计跨层数据收集机制,通过在路径发现阶段使用数据转发提高数据传输速率、降低延迟。文献[12]综合考虑了数据收集的延迟和容量问题。

此外,文献[13]首次研究如何尽可能传输较少的数据,同时传输的数据满足所有应用的要求。文献[14]首次提出在大规模无线传感器网络中,通过压缩采样理论收集数据,该方案能够降低整体通信负载且不会引入额外的计算。文献[15]分别研究了基于树和簇的数据收集和聚集协议。文献[16]分析了数据收集、聚集和选择的复杂度。文献[17]设计了一个自适应的数据收集近似算法。文献[18]提出了基于小波分段常值压缩的数据收集方法。文献[19]提出了一种分布式的高效节能的无线传感器网络数据收集协议。

同上述工作不同,本文研究的问题是考虑网络中节点出现故障以及阻塞等情况,如何收集网络中尽可能多的数据,并且数据收集的延迟较低。针对该问题,本文提出以森林作为路由结构进行数据收集的策略,森林表示的是具有多棵路由树的拓扑结构。

3 建立路由结构

在大多数的应用中,节点都是密集分布在监测区域内的,也即一个节点通常有多个邻居节点可以通信。为了避免由于信道竞争、节点故障等导致的大规模数据阻塞,本节提出建立多棵不相交的树形成森林进行数据收集。这里的不相交有2个方面的含义:1)网络中不存在某个节点在任意2棵路由树上使用除 sink 以外的相同节点作为父亲节点,即物理链路不相交;2)不存在某个节点在任意2棵路由树上使用相同的信道进行数据传输,也即逻辑链路不相交。

因此,对于一个待发送数据的节点,当一棵路由树上的父亲节点出现故障时,可以经由其他路由树上的父亲节点将数据发送到 sink。例如,给定拓扑结构如图 1(a)所示,图中的虚线表示节点之间可以进行通信。图 1(b)和图 1(c)是对图 1(a)给出的网络拓扑使用算法 1 建立的 2 棵不相交的路由树。从

图 1 中可以看到，当节点 1 因为出现故障而无法通信时，节点 6 和 7 可以通过第 2 棵路由树上的节点 2 进行数据传输，避免了节点 6、7、11 和 12 的数据的丢失。

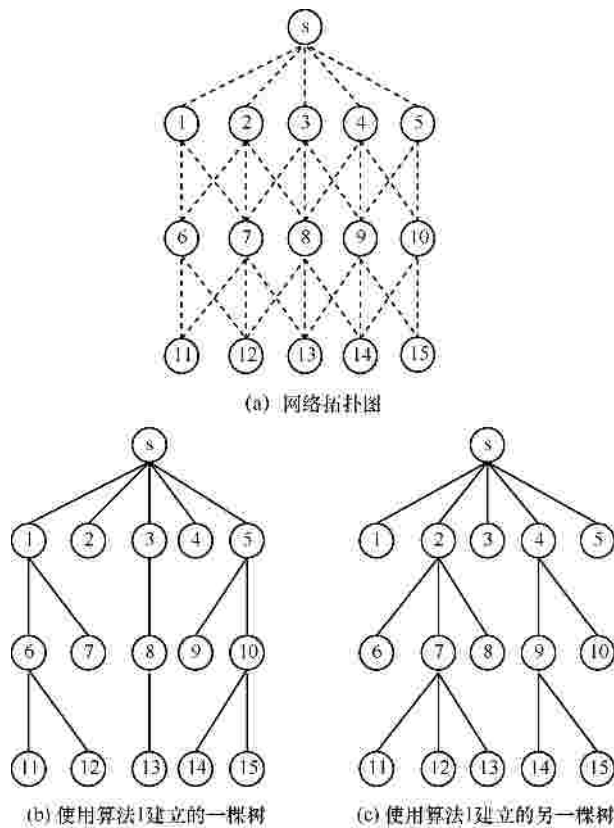


图 1 网络拓扑结构以及建立的 2 棵路由树

3.1 准备工作

考虑一个由 N 个传感器节点和一个汇聚节点 sink 组成的无线传感器网络，记为 $G=(V, E)$ ，其中， V 是网络中节点构成的集合， E 是网络中所有可能的通信链路集合。假设 sink 具有相对强大的能力，不会出现故障或者拥塞。假设网络中的每个节点配备 R 个 radio(无线收发器)，并且网络有 R 个可利用的正交信道，记为 $Channel=\{C_1, C_2, \dots, C_R\}$ 。设 r 和 ρ 分别表示节点配备 radio 的通信半径和干扰半径，在此假设所有 radio 具有相同的干扰半径和通信半径。设 hop_v 表示节点 v 距离 sink 的最短跳数， H 表示网络的高度，即网络中的节点距离 sink 的最大跳数。假设网络采用协议干扰模型，即 2 个节点可以成功通信当且仅当在接收节点的干扰范围内没有其他节点在同一时槽使用相同信道进行通信。

设 $P_v = \{u | u \in V, (u, v) \in E, hop_u < hop_v\}$ 表示节点 v 的候选父亲节点集合， N_{p_v} 表示节点 v 的候选父

亲节点的个数。那么，显然该网络可以构建出一个包含 X 棵不相交路由树的森林，其中，假设 T_1, T_2, L, T_X 是构建的 X 个节点，分别记做 s_1, s_2, L, s_X 。那么对于 sink 的任意邻居节点，扩展 v 的候选父亲节点集合为 $p_v = \{s_1, s_2, L, s_X\}$ ，设 $p_v^k (1 \leq k \leq X)$ 表示节点 v 在 k 棵路由树 T_k 上的父亲节点。令 $L_k = \{(v, p_v^k) | v \in V\}$ 表示集合 V 中的节点和它在第 k 棵路由树上的父亲节点形成的链路集合。

由于候选父亲节点集中的节点均是节点 v 的上层节点，因此，集合 L_k 中的链路一定能够形成一棵路由树，那么建立包含 X 棵路由树的森林的问题可以转化为：对于集合 V 中的每个节点 v 在其候选父亲节点集合 P_v 中选择 X 个不同的父亲节点的问题。

3.2 构造森林

设 F_v 表示在节点 v 的干扰半径内的节点构成的集合，即 $F_v = \{u | u \in V, u - v \leq r\}$ 。由协议干扰模型可知，在节点 v 接收数据时，集合 F_v 中的任意其他节点不能在同一时槽使用相同信道发送数据。而当节点 v 作为发送节点时，则 F_v 中的任意其他节点不能在同一时槽使用相同信道接收数据。例如，在图 1(a)中，节点 8 的干扰半径内的节点构成的集合为 $\{2, 3, 4, 12, 13, 14\}$ ，假设节点 8 在接收节点 12 发送的数据时，如果节点 2、3、4、13 和 14 中的任意节点使用同一信道发送数据则会由于信道干扰而传输失败。当节点 8 向节点 2 发送数据时，如果 3、4、12、13、14 中的任意节点使用同一信道接收其他节点发送的数据同样会发生冲突。

因此，在第 k 棵路由树上，与节点 v 同时传输数据会导致干扰的节点有 2 种类型：1) 以 F_v 中的任意节点作为父亲节点的节点 u ，即 $p_u^k \in F_v$ ；2) 在 p_v^k 的干扰范围内的任意节点，即 $F_{p_v^k}$ 中的节点，用 SI_v^k 表示节点 v 在第 k 棵树上的干扰节点集合，那么 $SI_v^k = F_{p_v^k} \cup \{u | p_u^k \in F_v\}$ 。对于 P_v 中的任意节点 u ，定义 $M(v, u, k)$ 表示节点 v 在第 k 棵树上选择节点 u 作为父亲节点的干扰度，如式(1)所示。

$$M(v, u, k) = |SI_v^k \cap \{F_u \cup \{w | p_w^k \in F_v\}\}| \quad (1)$$

显然，一个节点的候选父亲节点集合越大，在选择父亲节点时具有更多的可选择性，因此，按照节点的候选父亲节点个数由低到高为节点分配每棵路由树上的父亲节点。

基于森林的数据收集不会造成数据冗余，之所

以选择这种多棵树的路由结构，只是为了避免当一个节点出现问题时导致经过该节点的数据都无法传输成功，现在有多棵路由树，一个节点通信失败时，数据可以通过其他的路由树传输，而不是同时使用多条路由传输，因此不会产生数据冗余，也不会增大通信开销。

下面给出建立包含 X 棵树的森林的过程，详见算法 1。

1) 首先将集合 V 中的节点按照候选父亲节点个数 N_{p_v} 进行升序排列，然后从 $k=1$ 开始执行下述 2 个步骤，直至 $k=X$ 。

2) 按照候选父亲节点个数依次考虑每个节点 v ，对于节点 v 的候选父亲节点集合 P_v 中的任意节点 u ，计算选择 u 作为 v 在第 k 棵树上的父亲节点的干扰度 $M(v,u,k)$ ，选择干扰度 $M(v,u,k)$ 值最小的节点作为 v 的父亲节点 p_v^k 。在干扰度相同时，选择节点标号最小的节点作为 v 的父亲节点 p_v^k 。

3) 如果集合 V 中有节点未分配第 k 个父亲节点，则继续执行步骤 2)。否则更新其在第 k 棵路由树上的干扰节点集合 SI_v^k ，执行步骤 1)。

例如，给定网络拓扑结构如图 1(a)所示，使用算法 1 在建立第一棵路由树时，首先按照候选父亲节点集合的大小考虑，节点 6、10、11 和 15 都有 2 个候选父亲节点，是最少的，所以先考虑这 4 个节点，节点间的顺序则是按照节点编号顺序列出的，因此依次考虑节点 6、10、11 和 15，选择父亲节点加入到树中。考虑节点 7 有 3 个候选父亲节点，分别是 1、2 和 3。那么，如果节点 7 选择节点 1 作为父亲节点则与集合 {6} 干扰，因为节点 6 也在节点 1 的干扰半径内。如果选择 2 作为父亲节点则与集合 {6,8} 干扰。如果选择 3 作为父亲节点则与集合 {8,9} 干扰。因此，节点 7 选择节点 1 作为父亲节点。重复执行上述过程，直至所有节点都加入到树中，如图 1(b)所示。然后按照上述过程继续构建第 2 棵树，如图 1(c)所示。

算法 1 构造森林

输入：网络拓扑图 $G=(V,E)$

输出： X 棵路由树 T_1, T_2, \dots, T_X

1) 按照候选父亲节点个数 N_{p_v} 将集合 V 中的节点升序排列;

2) while $k \leq X$ do

3) for v in V do

4) for $u \in P_v$ do

5) 计算 $M(v,u,k)$ 的值;

6) end for;

7) 选择 $M(v,u,k)$ 值最小的节点作为 v 的父亲节点 p_v^k ; 在干扰度相同时，选择节点标号最小的节点作为 v 的父亲节点 p_v^k ;

8) end for

9) for v in V do
更新干扰节点集合 SI_v^k ;

end for

$k++$;

end while

3.3 分配信道

假设网络中每个节点的第 $i(1 \leq i \leq R)$ 个 radio 固定的工作在信道 C_i 上。设 T_1, T_2, \dots, T_X 为通过算法 1 建立的 X 棵路由树，集合 $L_k(1 \leq k \leq X)$ 表示路由树 T_k 上链路构成的集合。

令第 $k(1 \leq k \leq X)$ 棵路由树上的链路，即集合 L_k 中链路全部使用节点的第 k 个 radio 进行数据传输，即如果节点 v 以第 k 棵树作为路由树，那么节点 v 在通信时使用第 k 个 radio 通过信道 C_k 向父亲节点 p_v^k 发送数据。因此，任意 2 棵不同的路由树上的链路之间不存在信道干扰问题。

4 数据收集策略和分析

设 D 是网络拓扑图 G 的最大度，那么显然任意一棵路由树上的链路集合 $L_k(1 \leq k \leq X)$ 都可以划分为 D 个无冲突的链路子集进行通信。设 $S_i^k \subseteq L_k(1 \leq i \leq D, 1 \leq k \leq X)$ 表示第 k 棵路由树上链路集合的第 i 个子集，为方便描述，下面使用链路的发送节点来代表整条链路。

本节提出的数据收集策略是首先将森林中每棵路由树上的链路集合划分成 D 个无冲突的子集合，然后将 D 个时槽作为一个周期，在每个时槽调度 X 棵路由树上的固定链路子集，直到所有节点的数据都收集到 sink。因此，网络中的数据收集过程可以描述为重复执行下述步骤直到所有节点的数据都收集到 sink：设 t 是通信时槽，在时槽 t 调度链路集合 $S_i^1, S_i^2, \dots, S_i^X$ 中的链路进行数据传输，其中， $i=(t-1)\%D+1$ ，即对于 S_i^k 中的任意节点 v ， v 使用信道 C_k 向父亲节点 p_v^k 发送数据。对于节点 v ，如果 $v \in S_i^k$ ，则称 i 为节点 v 在第 k 棵路由树上的一个调度时间。

4.1 划分链路集合

在划分链路集合之前首先介绍调度时间差的定义。

定义 1 调度时间差。假设 t_1 和 t_2 是节点 v 的 2 个调度时间，那么 t_1 和 t_2 的时间差定义为 $t_1 \in t_2 = \min\{|t_1 - t_2|, D - |t_1 - t_2|\}$ 。

例如，假设 $D=15, t_1=1, t_2=14$ ，那么 t_1 和 t_2 的调度时间差为 2，表示节点 v 在 2 次调度之间的时间间隔为 2。设 $DV(v, k, i)$ 表示节点 v 在前 $k-1$ 棵路由树上的调度时间与 i 的最小差值，如式(2)所示。设 $Pro(v, k, i)$ 表示在第 k 棵路由树上以节点 v 作为发送节点的链路划分到第 i 个链路子集 S_i^k 的优先级，如式(3)所示。

$$DV(v, k, t) = \min\{t \in i | v \in S_i^j, j < k\} \quad (2)$$

$$Pro(v, k, i) = \frac{J}{hop_v} + DV(v, k, i) \quad (3)$$

$Pro(v, k, i)$ 由 2 个部分组成，1) 节点 v 到 sink 的距离，距离越近的节点优先级越高，其中， J 是一个大于 D 的参数。这是由于在数据收集过程中，越靠近 sink 节点传输的数据量越大，拥塞的概率越大，此外，这些节点也更可能会由于电池能量耗尽而出现故障。2) v 在前 $k-1$ 棵路由树上的调度时间与 i 的最小差值，差值越小的节点优先级越高。这是在无线传感器网络中，由于无线链路性能不稳定，链路可能会出现拥塞并持续一段时间。为减少传输失败的数据的等待时间，降低数据收集延迟，同一个节点在不同路由树上的传输数据的时槽之间的间隔应该尽可能小。

下面给出将路由树上的链路集合划分为 D 个子集合的过程。以路由树 $T_k(1 \leq k \leq X)$ 的第 $i(1 \leq i \leq D)$ 个链路集合 S_i^k 的划分为例进行说明。

1) 初始时令集合 S_i^k 为空集，首先将未加入前 $i-1$ 个调度集合的节点按照优先级降序排列，即将 $V - \bigcup_{1 \leq j < i} S_j^k$ 中的节点 v 按照 $Pro(v, k, t)$ 的值降序排列。

2) 依次考察每个节点 v ，如果对于集合 S_i^k 中的任意节点 u ，节点 v 与节点 u 在协议干扰模型下不冲突，即 $\|u - p_v^k\| > r$ 并且 $\|u - p_u^k\| > r$ ，则将节点 v 加入到集合 S_i^k 中。

3) 最后将前 $i-1$ 个调度集合 $\bigcup_{1 \leq j < i} S_j^k$ 的节点按照值降序排列。依次考察每个节点 v ，如果对于集合 S_i^k 中的任意节点 u ，节点 v 与节点 u 在协议干扰

模型下不冲突，即 $\|u - p_v^k\| > r$ 且 $\|u - p_u^k\| > r$ ，则将节点 v 加入到集合 S_i^k 中。

算法 2 划分链路集合

输入： X 棵路由树 T_1, T_2, \dots, T_X

输出： X 棵树的 D 个链路子集 $S_i^k(1 \leq k \leq X, 1 \leq i \leq D)$

- 1) $k=1$;
- 2) while $k \leq X$ do
- 3) $i=1$;
- 4) while $i < D$ do
- 5) S_i^k 置为空集，按照 $Pro(v, k, t)$ 的值将 $V - \bigcup_{1 \leq j < i} S_j^k$ 中的节点降序排列;
- 6) for 优先级最高的节点 v do
- 7) if 节点 v 与集合 S_i^k 中的任意节点 u 不冲突
- 8) then 将节点 v 加入到集合 S_i^k 中;
- 9) end for
- 10) 按 $Pro(v, k, t)$ 的值将 $\bigcup_{1 \leq j < i} S_j^k$ 中节点 v 降序排列;
- 11) for 优先级最高的节点 v do
- 12) if 节点 v 与集合 S_i^k 中的任意节点 u 不冲突
- 13) then 将节点 v 加入到集合 S_i^k 中;
- 14) end for
- 15) $i++$;
- 16) end while
- 17) $k++$;
- 18) end while

4.2 理论分析

下面对本文提出的基于森林的数据收集策略进行理论分析，并举例子进行说明。

定理 1 给定一个 N 个节点组成的传感器网络，高度为 H ，可以建立 X 棵不相交的树，设在一次数据收集中节点出现故障的平均概率是 a ，那么 sink 平均可以收集到 $N(1-a)(1-a^X)^{\frac{H-1}{2}}$ 个数据。

证明 假设对于网络中的任意节点 v 发送的数据分组，该数据分组能够成功传输到下一跳当且仅当节点 v 未出现故障，并且节点 v 在 X 棵树上的父亲节点集合至少有一个未出现故障，因此节点 v 将数据分组成功传输到下一跳的概率是 $(1-a)(1-a^X)$ 。由路由结构可知节点 v 到达 sink 的路径长度为 hop_v ，

因此节点 v 产生的数据分组成功传输到 sink 的概率为 $(1-a)(1-a^x)^{hop_v-1}$ 。那么，数据分组到达 sink 的平均概率为 $(1-a)(1-a^x)^{\frac{H}{2}-1}$ ，也即 sink 可以收集到 $N(1-a)(1-a^x)^{\frac{H}{2}-1}$ 个数据。

以一个例子进行说明，假设网络中有 100 个节点，网络高度为 12 层，节点的平均故障概率 a 为 0.05，每个节点传送一个数据分组的时间 t 为 1 个单位时间，节点传输数据失败，重传一个数据分组所需时间为 $1.2t$ ，并假设网络拓扑可以构造出包含 3 棵树的森林。那么，根据定理 1 中的公式 $N(1-a)(1-a^x)^{\frac{H}{2}-1}$ ，以 3 棵路由树收集数据时 sink 可以收集到大约 $100(1-0.05)(1-0.05^3)^{\frac{12}{2}-1} \approx 95$ 个数据分组，由于数据重传所产生的延迟为 $5 \times 1.2t = 6$ 个单位时间。而在以一棵树作为路由结构的数据收集中，sink 可以成功接收到的数据分组个数为 $100(1-0.05)(1-0.05^1)^{\frac{12}{2}-1} \approx 74$ ，由于数据重传所产生的延迟为 $26 \times 1.2t = 31.2$ 个单位时间。

5 模拟实验与结果分析

本文采用 Microsoft Visual C++ 6.0 编程环境模拟无线传感器网络，假设 400 个传感器节点均匀随机地分布在 $50\text{ m} \times 50\text{ m}$ 的监测区域内，sink 位于网络的中间位置，网络高度为 12 层，节点的通信半径和干扰半径均设置为 5 m。假设传感器网络的 MAC 层使用 TDMA 协议工作，即时间可以划分成若干个时槽。节点每个可利用的信道都有相同的带宽，设为 1。假设每个节点使用的任意 2 个不同的信道都是正交的，即每个节点在任意 2 个信道上的通信不存在干扰。设网络中节点出现故障的概率是 5%，sink 收集数据的成功率为 95%，则根据定理 1 可以得到森林中需要构建的树的棵数为 3。

假设每个节点每次产生一个数据，该数据可以在一个时槽内成功传输对于数据收集，所有节点在指定时间的所有感知数据集称为一个快照。收集一个快照的数据称为快照数据收集，收集多个连续快照的问题称为连续数据收集，数据收集延迟的单位是时槽。

下面分别进行节点的故障对收集到的数据量的影响实验以及链路拥塞对数据收集的延迟影响的实验。

5.1 数据收集量实验

下面对快照数据的收集进行实验，在该实验中分别对比森林中包含 2 棵路由树和 3 棵路由树时整个网络的分组丢失率，并且将其与基于连通支配集 (CDS) 的路由树^[3]进行对比。如图 2 所示为出现故障的节点个数对收集到的数据量的影响情况，其中，横坐标为出现故障的节点个数，在此设置为从 0~25 个，并且每隔 5 个做一次记录，这里的故障节点个数表示在一次数据收集过程中出现故障的节点个数，纵坐标表示收集到的数据个数。

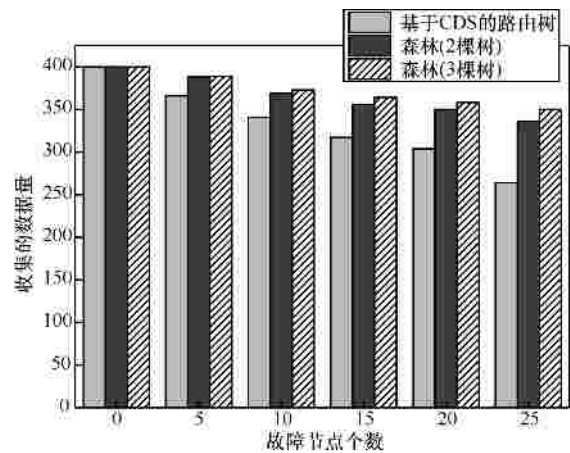


图 2 故障节点个数对数据收集量的影响

从图 2 中可以看出，随着出现故障的节点个数的不断增多，基于连通支配集的路由树收集到的数据量明显减少，大约增加 5 个故障节点收集到的数据量降低 60 个左右。这是由于在一棵路由树上的一个节点的故障不仅会导致自身数据丢失，还会导致其所有子孙节点的数据丢失。而 2 棵路由树和 3 棵路由树丢失的数据量较低，这是由于在以森林做路由结构的情况下，一个节点出现故障，其孩子节点可以转换到其他路由树上进行数据传输，从而保证未出现故障的节点在很大程度上有到达 sink 的路径。图 3 是连续数据收集中随着快照次数的增加 sink 收集到的数据量情况，也即收集到的数据个数随着网络使用时间的变化趋势。如图 3 所示，横坐标表示在连续数据收集中执行的快照次数，纵坐标表示 sink 收集到的数据个数。以森林作为路由结构收集到的数据量明显高于基于 CDS 的路由树收集到的数据量，并且随着快照次数的增多，这种优势愈加明显，在进行 40 次快照后包含 3 棵树的森林仍旧可以收集一半的数据量。此外，在初始的快照收集中包含 2 棵树的森林和包含 3 棵树的森林收集

到的数据量相差无几，但随着快照次数的增多，包含 3 棵树的森林优势逐渐明显，并且趋于稳定。这是由于初始时出现故障的节点个数较少，这些故障节点的子孙节点数据可以经由第 2 棵树传输到 sink，而不必使用第 3 棵树；当故障节点的个数随着快照次数增加时，网络中的一些节点必需使用第 3 棵树才能够传输数据。

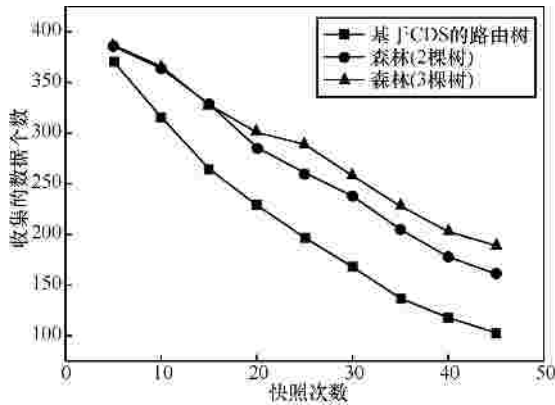


图 3 快照次数对数据收集量的影响

图 4 是节点出现故障的概率对收集的快照次数的影响实验。

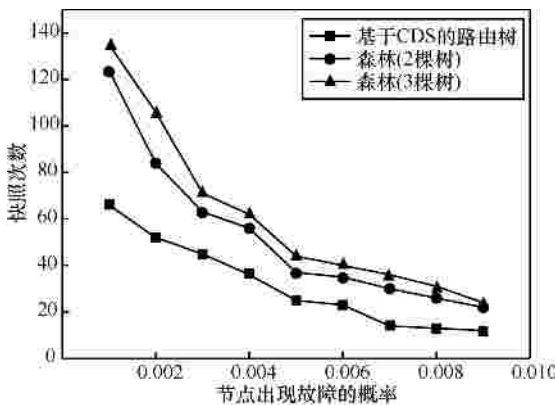


图 4 节点故障概率对快照次数的影响

如图 4 所示，横坐标是在数据收集过程中节点出现故障的概率，纵坐标表示在连续数据收集中收集到 200 以上数据的快照次数，即收集到网络中一半节点感知数据的快照次数。随着节点故障概率的增加，收集 200 个以上数据的快照次数随之减小，但是包含 3 棵树森林的快照次数总是基于 CDS 路由树的快照次数的 2 倍左右。

5.2 数据收集延迟实验

在链路拥塞的实验中，分别测试链路出现拥塞的概率以及拥塞的等待时间对数据收集的影响。拥

塞的等待时间表示在一条链路出现拥塞后该链路在一段时间内会一直处于拥塞状态，这段时间称为拥塞的等待时间。在实验中分别对采用 TAG 算法^[20]形成的单棵路由树和包含 3 棵路由树的森林数据收集的延迟进行了测试。为了实验的公平性，在 TAG 路由树数据收集过程中，令网络中的每个节点仍有 3 个可以使用的正交信道进行通信。

图 5 为链路出现拥塞的概率对数据收集延迟的影响。如图 5 所示，横坐标表示链路在每个调度时槽出现拥塞的平均概率为 0~0.2，设置拥塞等待时间为 3 个时槽，纵坐标表示数据收集的延迟。即图中可以看到随着链路拥塞概率的增大，收集延迟也随之增加，而以森林作路由结构的延迟要小于 TAG 的路由结构。这是由于在基于森林的数据收集过程中，在某一时刻一条链路出现拥塞后，该链路的发送节点在较短的时间内会以其他节点作为父亲节点重新传输失败的数据，降低了数据的等待时间。

图 6 所示为链路拥塞的等待时间对数据收集延迟的影响，在此设置每个时槽链路出现拥塞的概率为 0.05。横坐标为链路拥塞的等待时间，即 0~8 个时槽，纵坐标表示数据收集的延迟。从图 6 中可以看到，TAG 路由树的数据收集延迟几乎与等待时间成正比，而基于森林的数据收集延迟随着拥塞等待时间的增加增长相对缓慢。在 TAG 路由树的数据收集过程中，假设一条链路拥塞，如果在下一个调度时槽仍处于拥塞等待状态，那么该链路的发送节点只能等待下一个调度时槽的到来。而以森林作为路由结构的数据收集过程中，该链路的发送节点可以使用其他路由树的链路进行数据传输，而不必等待该链路恢复正常。

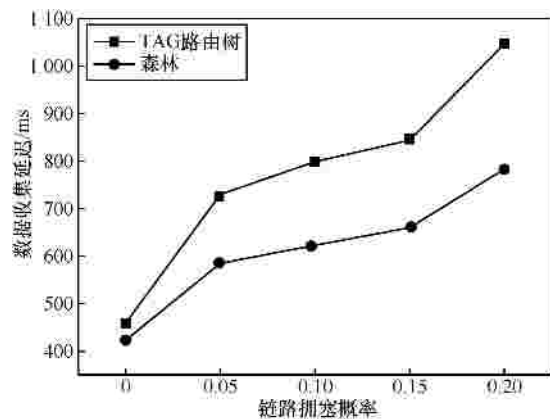


图 5 链路拥塞概率对延迟的影响

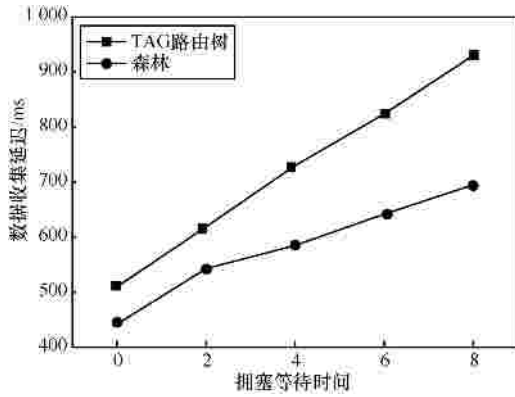


图6 拥塞等待时间对收集延迟的影响

6 结束语

在无线传感器网络中，基于 TAG 路由树的数据收集策略在节点出现故障时会造成较多的数据丢失，此外，节点的拥塞会造成较高的数据收集延迟。针对上述问题，本文提出建立森林作为收集网络中数据的路由结构，并且设计了基于森林的数据收集算法。理论分析和实验结果表明，在网络中节点出现故障概率较高或者拥塞严重的情况下，本文提出的方法能以较低的延迟收集到网络中的大部分数据。

参考文献：

- [1] 李凤保, 李凌. 无线传感器网络技术综述[J]. 仪器仪表学报, 2005, 26(8): 559-561.
LI F B, LI L. Survey on wireless sensor network techniques[J]. Chinese Journal of Scientific Instrument, 2005, 26(8): 559-561.
- [2] CHEN S, HUANG M, TANG S, et al. Capacity of data collection in arbitrary wireless sensor networks[J]. Parallel and Distributed Systems, 2012, 23(1): 52-60.
- [3] JI S, LI Y, JIA X. Capacity of dual-radio multi-channel wireless sensor networks for continuous data collection[C]/INFOCOM, 2011 Proceedings IEEE. c2011: 1062-1070.
- [4] JI S, BEYAH R, CAI Z. Snapshot/continuous data collection capacity for large-scale probabilistic wireless sensor networks[C]/INFOCOM 2012 Proceedings IEEE. c2012: 1035-1043.
- [5] CHEN S, WANG Y, LI M, et al. Data collection capacity of random-deployed wireless sensor networks[C] //Global Telecommunications Conference, 2009. IEEE. c2009: 1-6.
- [6] JI S, CAI Z. Distributed data collection and its capacity in asynchronous wireless sensor networks[C] //INFOCOM, 2012 Proceedings IEEE. c2012: 2113-2121.
- [7] CHENG C T, TSE C K, LAU F C M. A delay-aware data collection network structure for wireless sensor networks[J]. Sensors Journal, 2011, 11(3): 699-710.
- [8] INCEL O D, GHOSH A, KRISHNAMACHARI B, et al. Fast data collection in tree-based wireless sensor networks[J]. IEEE Transactions on Mobile Computing, 2012, 11(1): 86-99.
- [9] INCEL O D, GHOSH A, KRISHNAMACHARI B. Scheduling algorithms for tree-based data collection in wireless sensor networks[M]//Theoretical Aspects of Distributed Computing in Sensor Networks. Springer Berlin Heidelberg, 2011: 407-445.
- [10] SEKSAN L, EDWARD J, COYLE. Optimizing the collection of local decisions for time-constrained distributed detection in WSNs[C]//

INFOCOM, 2013 Proceedings IEEE. c2013: 1923-1931.

- [11] HUANG C, LIN T, CHEN L, et al. XD: a cross-layer designed data collection mechanism for mission-critical WSN in urban buildings[C]//International Conference on Mobile Data Management: Systems, Services and Middleware 2009.
- [12] CHEN S, WANG Y, LI M, et al. Order-optimal data collection in wireless sensor networks: delay and capacity[C]//6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. c2009: 1-9.
- [13] FANG X, GAO H, LI J, et al. Application-aware data collection in Wireless Sensor Networks[C]/INFOCOM, 2013 Proceedings IEEE. c2013: 1645-1653.
- [14] LUO C, WU F, SUN J, et al. Compressive data gathering in large-scale wireless sensor networks[C]//The 15th Annual International Conference on Mobile Computing and Networking. ACM, c2009: 145-156.
- [15] WANG W, WANG B, LIU Z, et al. A cluster-based and tree-based power efficient data collection and aggregation protocol for wireless sensor networks[J]. Information Technology Journal, 2011, 10(3): 557-564.
- [16] LI M, WANG Y, WANG Y. Complexity of data collection, aggregation, and selection for wireless sensor networks[J]. IEEE Transactions on Computers, 2011, 60(3): 386-399.
- [17] WANG C, MA H, HE Y, et al. Adaptive approximate data collection for wireless sensor networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(6): 1004-1016.
- [18] 李杨, 郭龙江, 李金宝, 等. 传感器网络基于小波分段常值压缩的数据收集研究[J]. 仪器仪表学报, 2013, 34(1): 119-127.
LI Y, GUO L J, LI J B, et al. Data collection using wavelet-segment constant compression in wireless sensor networks[J]. Chinese Journal of Scientific Instrument, 2013, 34(1): 119-127.
- [19] 史久根, 胡小博. 高效节能的无线传感器网络数据收集协议[J]. 电子测量与仪器学报, 2012, 26(5): 437-445.
SHI J G, HU X B. Energy-efficient data gathering protocol for wireless sensor networks[J]. Journal of Electronic Measurement and Instrument, 2012, 26(5): 437-445.
- [20] MADDEN S, FRANKLIN M J, HELLERSTEIN J M, et al. Tag: a tiny aggregation service for ad hoc sensor networks[J]. OSDI Conf, 2002, 36(1): 1-28.

作者简介：



张伟平 (1964-), 女, 黑龙江哈尔滨人, 黑龙江大学工程师, 主要研究方向为无线传感器网络。

郭亚红 (1972-), 女, 黑龙江双鸭山人, 黑龙江大学副教授, 主要研究方向为无线传感器网络。

王蒙 (1989-), 女, 黑龙江牡丹江人, 黑龙江大学硕士生, 主要研究方向为无线传感器网络。

倪林雨 (1990-), 男, 黑龙江庆安人, 黑龙江大学硕士生, 主要研究方向为无线传感器网络。

李金宝 (1969-), 男, 黑龙江庆安人, 博士, 黑龙江大学教授, 主要研究方向为无线传感器网络、数据库原理、移动计算和并行计算。