

基于动态最小生成树路由协议的数据聚融算法*

彭海霞[†] 赵海 李大舟 林川

(东北大学信息科学与工程学院, 沈阳 110819)

(2013年12月13日收到; 2014年1月20日收到修改稿)

大规模的地理环境监测, 以及用来传输与处理数据的物理基础设施无法和监测区域的规模保持同样的增速, 使得不可靠链路下的数据采集与处理呈现出一种饱和和流状态, 无线传感器网络的能力看似难以稳定. 另一方面, 尽管理想网络模型的计算结果足够精确, 然而, 由于与实际应用偏差甚大, 使得网络用户无法充分地分析和利用从工业现场所获得到的网络感知数据, 并且没有针对网络规模和性能对数据聚融的影响进行分析. 为此, 本文提出以“过渡区”作为工业现场仿真的假设条件, 并在此基础上提出了一种面向实际应用的数据聚融算法, 即基于可信度的数据聚融算法(R算法). 在具体设计R算法过程中, 选用聚集和操作符SUM为例, 通过对网络提供出的近似聚集和加以自动分析、综合, 针对相对误差界限 ϵ , 计算出近似聚集和的可信度的下限 η ; 并将近似聚集和、参数 η 一同提供给用户, 在为用户提供网络概要信息的同时, 还提供了参数 η 作为对信息可信度的判断, 以便指导用户对数据聚集结果进行深度处理和提高网络的感知性能. 仿真实验描述了过渡区内由于信噪比导致的链路不可靠所引起的 η 的变化规律; 讨论了网络性能和规模对 η 的影响, 随着网络运行周期的增加和网络规模的增大, η 的值将逐渐靠近0; 从而为WSNs从理论模型投入实际工业应用提供了理论依据和经验公式.

关键词: 无线传感器网络, 数据聚融, DMSTRP, 可信度

PACS: 02.50.-r, 02.50.Sk, 05.60.-k, 06.20.Dk

DOI: 10.7498/aps.63.090206

1 引言

近年来, 微电子技术、无线通信技术以及传感器技术的快速发展, 使得低成本、低功耗、多功能的无线传感器节点成为可能^[1]. 与此同时, 无孔不入的监视能力亦使得传感器网络被广泛应用于众多领域, 如栖息地的监测^[2,3], 对象跟踪^[4,5], 环境监测^[6-8], 国防军事^[9,10], 灾害管理^[11], 健康监测^[12-14], 以及智能环境^[15]等等. 上述应用中, 数据聚集和数据融合在为用户提供网络概要信息的过程中起到了十分重要的作用. 为促进无线传感器网络(wireless sensor networks, WSNs)的应用发展, 有关数据聚集和数据融合的研究被陆续展开^[16-21]. 然而, 多数研究的目的在于探讨理想网

络模型的网络容量问题, 几乎没有涉及工业现场的实际应用情况, 而且很少考虑传感器节点(sensor nodes, SNs)由于低电池供电所引起的资源限制. 此外, 数据聚集和数据融合的不同侧重点(数据聚集侧重于数据的传输过程, 而数据融合的侧重点在于感知数据的分析与融合)对WSN是同等重要的, 而已有的研究很少对两者进行同时考虑, 根本没有针对网络性能和网络规模对数据聚融的影响进行分析.

从多数文献来看, 针对WSNs的数据聚集和数据融合算法的研究, 大多是建立在确定的网络模型之上, 且忽略了网络中SNs的能量受限问题. 在确定性网络中, 两个节点的连接是确定性的, 即在没有碰撞下, 感知数据在该链路上将成功传输. 若两个节点断开, 则假定它们之间无法直接通信. 然而,

* 国家科技支撑计划(批准号: 2012BAH82F04)、国家自然科学基金(批准号: 61101121)和国家高技术研究发展计划(863计划)(批准号: 2013AA102505)资助的课题.

[†] 通讯作者. E-mail: penghaixia@neuera.com

在实际工业应用中,通常不能对网络中SNs的电池进行充电,即SNs是能量受限的;而且网络的链路并非确定性连接,其在确定性连接与断开之间还存在一个以一定概率来成功传送数据的区段,使得网络中约90%以上的链路变得不可靠,这种链路被称为有损链路.即使没有碰撞,有损链路上的数据也是以一定的概率传输成功,由有损链路构成的WSNs称为概率型网络.针对上述问题,本文选择基于簇的WSN模型^[22]以提高网络的能源利用率,并引入“过渡区”^[23]来仿真实际工业应用中的链路不可靠问题.

对WSNs,数据聚集是指SNs收集监测区域的物理信息,并将信息向Sink传递^[24],主要完成信息感知和数据传输两项任务;数据融合则是在一定准则下对感知数据加以自动分析、综合,以完成所需的决策和评估任务^[25],侧重于对感知数据的融合分析.数据传输和融合分析对WSNs同等重要,而已有的研究很少对两者进行同时考虑.为此,本文在数据聚集和数据融合的基础上提出了数据聚融的概念,依据用户的不同需求,在对WSNs中感知数据进行采集、传输、汇聚、转发后,Sink将对接收到的汇聚结果进行分析融合.即数据聚融是在网络采集到感知数据的基础上,使聚集而得的感知数据发挥更大的作用,以辅助人们进行态势/环境判定、规划、探测、验证、诊断.

为具体研究WSNs中的数据聚融算法,并对网络性能和规模对数据聚融结果的影响进行分析,本文提出了基于可信度的数据聚融算法(data fusion algorithm based on reliability),简称为R算法.在R算法中,为方便网络用户对数据聚集结果的进一步处理,定义了用于评估聚集结果可信度的参数 η ,即可信度的下限.通过计算近似聚集结果的 η ,并将网络的近似聚集结果和参数 η 一同传送到用户,达到指导用户对近似聚集结果进行深度处理(是将聚集结果作为决策依据,又或参考依据,还是直接丢弃此次的聚集结果并再次聚集)和提高网络感知性能的目的.

2 网络模型

假定一个概率型WSN中, t 时刻的活动节点总数为 N_t , $\inf(N_t)$ 为 N_t 的下限值.在网络运行期间, N_t 随时间变化,且对Sink而言 N_t 的值是未知的.令 s_{ti} ($1 \leq i \leq N_t$)表示节点 i 在时刻 t 的感知

数据,则 $S_t = \{s_{t1}, s_{t2}, \dots, s_{tN_t}\}$ 为时刻 t 所有感知数据的集合,并假定 s_{ti} ($1 \leq i \leq N_t$)保存在活动节点 i 上.由于网络中的感知数据是有界的,在此设 $\sup(S_t)$ 为所有感知数据的上限值.

本文选用基于簇的WSN模型,利用动态最小生成树路由协议(dynamic minimal spanning tree routing protocol, DMSTRP)^[26]将网络的监测区域划分成若干个相互独立的簇,其簇的形成过程同于文献^[27].设以CH为父节点的节点为第一层节点,以第一层节点为父节点的节点为第二层节点,依此类推.假定网络的整个监测区域均被活动节点所覆盖,针对时刻 t ,网络被划分成 n 个互不重叠的簇,且 ω 为所有簇中节点层数的最大值.设 m_{t1} 为第一层节点的节点总数, $m_{t2}, m_{t3}, \dots, m_{t\omega}$ 分别为第二层节点的节点总数,第三层节点的节点总数,……,第 ω 层节点的节点总数;并设 $S_{1t} = \{s_{1t1}, s_{1t2}, s_{1t3}, \dots, s_{1tm_{t1}}\}$ 为第一层节点的感知数据的集合, $S_{2t}, S_{3t}, \dots, S_{\omega t}$ 分别为第二层,第三层,……,第 ω 层节点的感知数据的集合.其中, $S_t, S_{1t}, S_{2t}, S_{3t}, \dots, S_{\omega t}$ 间的关系为

$$S_{1t} \subseteq S_t, S_{2t} \subseteq S_t, S_{3t} \subseteq S_t, \dots, S_{\omega t} \subseteq S_t, \\ S_t = \{S_{1t}, S_{2t}, S_{3t}, \dots, S_{\omega t}\}.$$

假定网络中的节点是均匀分布的.当Sink节点在时刻 t 发起一次数据采集时,令 d 为网络中任意两相连节点间的距离,链路上的数据以概率 q 传输成功.CH对数据的转发采用超时重传,即感知数据到达CH后均可保证成功到达Sink.由此可知,网络中第一层节点的感知数据将以概率 q 被成功传送到Sink,第二层节点、第三层节点……第 ω 层节点的感知数据被成功传送到Sink的概率将分别为 $q^2, q^3, \dots, q^\omega$.针对时刻 t ,令 $B^{(q)}$ 为Sink中来自第一层节点的感知数据的集合, $B^{(q^2)}, B^{(q^3)}, \dots, B^{(q^\omega)}$ 分别为来自第二层节点、第三层节点……第 ω 层节点的感知数据的集合.概率 q 与距离 d 间的关系将在下一小节中给出.

具体研究过程中,以“聚集和”(SUM)操作符为例,对基于可信度的数据聚融算法进行了设计和分析.针对时刻 t ,监测区域内的精确SUM为

$$\text{Sum}(S_t) = \sum_{i=1}^{m_{t1}} S_{1ti} + \sum_{i=1}^{m_{t2}} S_{2ti} + \sum_{i=1}^{m_{t3}} S_{3ti} + \dots \\ + \sum_{i=1}^{m_{t\omega}} S_{\omega ti},$$

下面将给出 ε 估计、可信度、 η 、无偏估计的定义. 其中 ε 为相对误差界限, 此处指近似聚集和与精确聚集和间的先对误差界限.

定义 1 (ε 估计) 对于给定的 ε ($\varepsilon \geq 0$), 近似聚集结果 \hat{I}_t (部分感知数据参与到聚集操作时的结果) 称之为精确结果 I_t (所有感知数据参与到聚集操作时的结果) 的 ε 估计当且仅当

$$\left| \frac{\hat{I}_t - I_t}{I_t} \right| \leq \varepsilon.$$

定义 2 (可信度) 对于给定的网络和 ε ($\varepsilon \geq 0$), 可信度表示近似聚集结果 \hat{I}_t 为精确聚集结果 I_t 的 ε 估计的概率.

定义 3 (η) η 表示近似聚集结果 \hat{I}_t 为精确聚集结果 I_t 的 ε 估计的概率的最小值, 即存在

$$\eta \leq P \left(\left| \frac{\hat{I}_t - I_t}{I_t} \right| \leq \varepsilon \right),$$

称 η 为可信度的下限.

定义 4 (无偏估计) \hat{I}_t 称为 I_t 的无偏估计当且仅当 \hat{I}_t 的数学期望等于 I_t , 即存在 $E(\hat{I}_t) = I_t$.

参数 η 的求解将在 R 算法中完成, 其中 R 算法的主要步骤如下:

第一步 Sink 发起一次数据采集, 并将数据采集命令传送至各 CH, CH 转发采集命令至其所有子节点.

第二步 依据文献 [26], 接收到采集命令的普通节点将其感知数据传送至 CH, CH 对接收到的感知数据进行分析处理后, 再将处理结果经由其他 CH 上传至 Sink.

第三步 依据公式 (理论推导见第三节)

$$\begin{aligned} \text{Sum}(S_t) &= \frac{1}{q} \sum_{S_{1ti} \in B(q)} S_{1ti} \\ &+ \frac{1}{q^2} \sum_{S_{2ti} \in B(q^2)} S_{2ti} + \dots \\ &+ \frac{1}{q^\omega} \sum_{S_{\omega ti} \in B(q^\omega)} S_{\omega ti}, \end{aligned}$$

Sink 对接收到的感知数据处理后得到“近似聚集和” $\text{Sum}(S_t)$.

第四步 Sink 依据 η 的求解公式

$$\varphi_{\delta/2}^2 \leq \frac{q^\omega \inf(N_t) \inf(S_t) \varepsilon^2}{(1 - q^\omega) \sup(S_t)}$$

(推导过程见第 4 节) 得到 η 的值, 并将 η 和 $\text{Sum}(S_t)$ 一同传送至网络用户, 至此, 算法结束.

结合上述对 R 算法的介绍可知, 对 R 算法而言, 其关键在于 η 的求解. η 的求解问题可描述如下:

输入 1) $S_{1t} = \{s_{1t1}, s_{1t2}, s_{1t3}, \dots, s_{1tm_{t1}}\}$, $S_{2t}, S_{3t}, \dots, S_{\omega t}$.

2) ε ($\varepsilon \geq 0$), ω , 和 q .

3) SUM 操作符.

输出 “近似聚集和” $\text{Sum}(S_t)$ 的可信度的下限 η .

3 R 算法的数学原理

结合前文的分析可知, R 算法是在网络近似聚集结果可以替代精确结果的前提下建立起来的, 而如何获取链路成功传送感知数据的概率 q 则是实现 R 算法的基础. 为此, 本节对“聚集和”的估计以及概率值 q 的求解公式进行了分析.

首先是针对“聚集和”估计器进行研究. 定义 SUM 的估计器为 $\text{Sum}(S_t)$, 结合第 2 节中相关变量的物理含义, 感知数据和的数学估计器可计算如下:

$$\begin{aligned} \text{Sum}(S_t) &= \frac{1}{q} \sum_{S_{1ti} \in B(q)} S_{1ti} \\ &+ \frac{1}{q^2} \sum_{S_{2ti} \in B(q^2)} S_{2ti} + \dots \\ &+ \frac{1}{q^\omega} \sum_{S_{\omega ti} \in B(q^\omega)} S_{\omega ti}. \end{aligned} \quad (1)$$

结合无偏估计的定义, 定理 1 将证明“近似聚集和” $\text{Sum}(S_t)$ 是“精确聚集和” $\text{Sum}(S_t)$ 的无偏估计.

定理 1 设 $\text{Sum}(S_t)$ 的数学期望为 $E(\text{Sum}(S_t))$, 方差为 $\text{Var}(\text{Sum}(S_t))$, 那么 $E(\text{Sum}(S_t))$ 与 $\text{Var}(\text{Sum}(S_t))$ 分别满足

$$E(\text{Sum}(S_t)) = \text{Sum}(S_t), \quad (2)$$

$$\text{Var}(\text{Sum}(S_t)) \leq \sup(S_t) \text{Sum}(S_t) \frac{1 - q^\omega}{q^\omega}. \quad (3)$$

证明 对于任意的 i ($1 \leq i \leq m_{tk}$), 设随机变量 $X_{1i}, X_{2i}, X_{3i}, \dots, X_{\omega i}$ 分别满足

$$X_{1i} = \begin{cases} 1, & S_{1ti} \in B(q), \\ 0, & S_{1ti} \notin B(q), \end{cases} \quad (4)$$

...

$$X_{\omega i} = \begin{cases} 1, & S_{\omega ti} \in B(q^\omega), \\ 0, & S_{\omega ti} \notin B(q^\omega), \end{cases} \quad (5)$$

显然存在,

$$\begin{aligned} P(X_{1i} = 1) &= q, & P(X_{1i} = 0) &= 1 - q, \\ P(X_{2i} = 1) &= q^2, & P(X_{2i} = 0) &= 1 - q^2, \dots, \\ P(X_{\omega i} = 1) &= q^\omega, & P(X_{\omega i} = 0) &= 1 - q^\omega. \end{aligned}$$

参考文献 [28] 可知, 存在 (2) 式成立. 同时, 依据前文分析, 对于 $1 \leq i \neq j \leq m_{tk}$, 随机变量 X_{ki} 与 X_{kj} 是相互独立的. 由 X_{ki} ($1 \leq k \leq \omega$) 的分布可得, 存在

$$\begin{aligned} \text{Var}(X_{1i}) &= q(1 - q), \\ \text{Var}(X_{2i}) &= q^2(1 - q^2), \dots, \\ \text{Var}(X_{\omega i}) &= q^\omega(1 - q^\omega). \end{aligned}$$

由于两相互独立事件 X 和 Y , 其方差存在

$$D(X + Y) = D(X) + D(Y)$$

的关系. 上述分析中, 结合

$$\begin{aligned} \text{Var}(X_{1i}) &= q(1 - q), \dots, \\ \text{Var}(X_{\omega i}) &= q^\omega(1 - q^\omega), \end{aligned}$$

可知

$$\begin{aligned} &\text{Var}(\text{Sum}(S_t)) \\ &= \frac{(1 - q)}{q} \sum_{i=1}^{m_{t1}} S_{1ti}^2 + \frac{(1 - q^2)}{q^2} \sum_{i=1}^{m_{t2}} S_{2ti}^2 + \dots \\ &+ \frac{(1 - q^\omega)}{q^\omega} \sum_{i=1}^{m_{t\omega}} S_{\omega ti}^2 \leq \frac{(1 - q)}{q} \sup(S_t) \sum_{i=1}^{m_{t1}} S_{1ti} \\ &+ \frac{(1 - q^2)}{q^2} \sup(S_t) \sum_{i=1}^{m_{t2}} S_{2ti} + \dots \\ &+ \frac{(1 - q^\omega)}{q^\omega} \sup(S_t) \sum_{i=1}^{m_{t\omega}} S_{\omega ti} = \frac{(1 - q^\omega)}{q^\omega} \sup(S_t) \\ &\times \left[\frac{q^{\omega-1}}{(1 + q + q^2 + \dots + q^{\omega-1})} \sum_{i=1}^{m_{t1}} S_{1ti} \right. \\ &+ \frac{q^{\omega-2}}{(1 + q^2 + q^4 + \dots + q^{\omega-2})} \sum_{i=1}^{m_{t2}} S_{2ti} \\ &+ \frac{q^{\omega-3}}{(1 + q^3 + q^6 + \dots + q^{\omega-3})} \sum_{i=1}^{m_{t3}} S_{3ti} \\ &\left. + \dots + \sum_{i=1}^{m_{t\omega}} S_{\omega ti} \right]. \end{aligned} \tag{6}$$

结合 (6) 式及

$$\begin{aligned} \frac{q^{\omega-1}}{(1 + q + q^2 + \dots + q^{\omega-1})} &\leq 1, \\ \frac{q^{\omega-2}}{(1 + q^2 + q^4 + \dots + q^{\omega-2})} &\leq 1, \end{aligned}$$

$$\frac{q^{\omega-3}}{(1 + q^3 + q^6 + \dots + q^{\omega-3})} \leq 1, \dots$$

可得

$$\begin{aligned} &\text{Var}(\text{Sum}(S_t)) \\ &\leq \frac{(1 - q^\omega)}{q^\omega} \sup(S_t) \times \left[\sum_{i=1}^{m_{t1}} S_{1ti} + \sum_{i=1}^{m_{t2}} S_{2ti} \right. \\ &\quad \left. + \sum_{i=1}^{m_{t3}} S_{3ti} + \dots + \sum_{i=1}^{m_{t\omega}} S_{\omega ti} \right] \\ &= \frac{(1 - q^\omega)}{q^\omega} \sup(S_t) \times \text{Sum}(S_t). \end{aligned} \tag{7}$$

定理 1 表明, “近似聚集和” $\text{Sum}(S_t)$ 是“精确聚集和” $\text{Sum}(S_t)$ 的无偏估计, 且存在 $\text{Sum}(S_t)$ 的方差 $\text{Var}(\text{Sum}(S_t))$, 其上限值与 q 成反比, 即 $\text{Var}(\text{Sum}(S_t))$ 的上限随着 q 的增大可以任意地小. 因而, 参考文献 [29] 可知, 随着概率 q 的增大, $\text{Sum}(S_t)$ 与“精确聚集和” $\text{Sum}(S_t)$ 之间的相对误差将逐渐减小, 只要 q 足够大, 两者间的相对误差就可以达到任意小. 即用“近似聚集和”来替代“精确聚集和”是具有一定的数学合理性.

下面我们将探讨 q 值的计算方法. 已有的研究表明, 概率值 q 是距离 d (任意两相连节点间链路的距离) 的函数. 参考文献 [30] 和文献 [31], 得到

$$\begin{aligned} q &= \left\{ 1 - 0.5 \times \exp \left[- \frac{1}{1.28} (P_{\text{trans}} - PL(d_0)) \right. \right. \\ &\quad \left. \left. - 10\beta \log_{10} \left(\frac{d}{d_0} \right) - X_\sigma - P_n \right] \right\}^{8f}. \end{aligned} \tag{8}$$

文献 [31] 将信道中噪声对信号的影响映射到链路距离上, 不同的链路距离将导致信道的信噪比存在不同, 最终使得链路成功传输数据的概率 q 不同. (8) 式中, P_{trans} 为发送端的输出功率, $PL(d)$ 为特定位置的路径损耗, d_0 为参考距离, β 为路径损耗指数, X_σ 为均值为 0 的高斯分布随机变量, P_n 为网络的平台噪声, f 表示帧的长度.

4 η 的求解

依据前文对 R 算法的介绍可知, 在 R 算法的 4 个步骤中, 其关键在于 $\text{Sum}(S_t)$ 的可信度下限 η 的求解. 本文所选用的 η 的求解方法中, 主要包括以下三步:

- 1) 证明 $\text{Sum}(S_t)$ 服从正态分布;
- 2) 将正态分布转化为标准正态分布;

3) 利用标准正态分布的相关特性来求解 η 的值.

首先, 我们需要证明 $\text{Sum}(S_t)$ 服从正态分布. 对于任意 i , 令变量 Y_{ki} ($1 \leq K \leq \omega$) 满足如下公式:

$$Y_{1i} = \begin{cases} S_{1ti}, & S_{1ti} \in B^{(q)}, \\ 0, & S_{1ti} \notin B^{(q)}, \end{cases} \quad (9)$$

.....

$$Y_{\omega i} = \begin{cases} S_{\omega ti}, & S_{\omega ti} \in B^{(q^\omega)}, \\ 0, & S_{\omega ti} \notin B^{(q^\omega)}. \end{cases} \quad (10)$$

第3节的分析表明,

$$\begin{aligned} \text{Sum}(S_t) &= \frac{1}{q} \sum_{S_{1ti} \in B^{(q)}} S_{1ti} \\ &+ \frac{1}{q^2} \sum_{S_{2ti} \in B^{(q^2)}} S_{2ti} + \dots \\ &+ \frac{1}{q^\omega} \sum_{S_{\omega ti} \in B^{(q^\omega)}} S_{\omega ti}. \end{aligned}$$

考虑到 $\text{Sum}(S_t)$ 为各层节点的感知数据的总和, 且 N 个相互独立的正态分布的线性组合仍服从正态分布. 又文献 [32] 表明, 对网络中的任意层而言, 若该层节点的感知数据满足 Lyapunov 条件, 则该层节点的感知数据之和将服从正态分布. 为此, 定理2将证明网络中第一层的感知数据、第二层的感知数据、.....、第 ω 层的感知数据分别满足 Lyapunov 条件, 即各层节点的感知数据之和均服从正态分布, 故而有 $\text{Sum}(S_t)$ 符从正态分布.

定理2 网络中 ω 组随机变量序列 Y_{ki} ($1 \leq K \leq \omega$) 满足 Lyapunov 条件, 即 $\exists \xi_k > 0$ 满足如下公式:

$$\lim_{m_{tk} \rightarrow \infty} \frac{1}{s_{m_{tk}}^{2+\xi_k}} \sum_{i=1}^{m_{tk}} E(|Y_{ki} - \mu_{ki}|^{2+\xi_k}) = 0, \quad (11)$$

其中, $1 \leq K \leq \omega$, m_{tk} 为 t 时刻第 k 层节点的总数,

$$s_{m_{tk}}^2 = \sum_{i=1}^{m_{tk}} \sigma_{ki},$$

且对于 $\forall i$ ($1 \leq i \leq m_{tk}$), 存在 $\mu_{ki} = E(Y_{ki})$ 和 $\sigma_{ki} = \text{Var}(Y_{ki})$.

证明 结合文章第2节的内容以及相关变量的物理含义可知:

第1层节点 ($k = 1$):

针对第一层节点, 存在

$$\begin{aligned} \mu_{1i} &= E(Y_{1i}) = qS_{1ti}, \\ \sigma_{1i} &= \text{Var}(Y_{1i}) = s_{1ti}^2 q(1-q). \end{aligned}$$

令 $\xi_1 = 1$. 依据上述分析可知, 对于 $\forall i$ ($1 \leq i \leq m_{t1}$), 有

$$\begin{aligned} E(|Y_{1i} - \mu_{1i}|^{2+\xi_1}) &= E(|Y_{1i} - \mu_{1i}|^3) \\ &= q(s_{1ti} - qs_{1ti})^3 + (1-q)(qs_{1ti})^3 \\ &= s_{1ti}^3 q(1-q)(1-2q+2q^2), \end{aligned} \quad (12)$$

$$\begin{aligned} s_{m_{t1}}^{2+\xi} &= s_{m_{t1}}^3 = \sum_{i=1}^{m_{t1}} s_{1ti}^2 q(1-q) \\ &\times \sqrt{\sum_{i=1}^{m_{t1}} s_{1ti}^2 q(1-q)}. \end{aligned} \quad (13)$$

结合 (12) 和 (13) 式, 可得

$$\begin{aligned} \lim_{m_{t1} \rightarrow \infty} \frac{1}{s_{m_{t1}}^3} \sum_{i=1}^{m_{t1}} E(|Y_{1i} - \mu_{1i}|^3) &= \frac{1-2q+2q^2}{\sqrt{q(1-q)}} \lim_{m_{t1} \rightarrow \infty} \frac{\sum_{i=1}^{m_{t1}} s_{1ti}^3}{\sum_{i=1}^{m_{t1}} s_{1ti}^2 \sqrt{\sum_{i=1}^{m_{t1}} s_{1ti}^2}} \\ &\leq \frac{1-2q+2q^2}{\sqrt{q(1-q)}} \frac{\sup(S_{1t})^3}{\inf(S_{1t})^3} \lim_{m_{t1} \rightarrow \infty} \frac{1}{\sqrt{m_{t1}}}, \end{aligned} \quad (14)$$

其中, $\inf(S_{1t})$ 和 $\sup(S_{1t})$ 分别表示第一层节点的感知数据的最小值和最大值, 由于 $|\inf(S_{1t})|, |\sup(S_{1t})| \ll +\infty$, 因此存在

$$\lim_{m_{t1} \rightarrow \infty} \frac{1}{s_{m_{t1}}^3} \sum_{i=1}^{m_{t1}} E(|Y_{1i} - \mu_{1i}|^3) \leq 0.$$

又 $s_{m_{t1}}^3 \geq 0$, $E(|Y_{1i} - \mu_{1i}|^3) \geq 0$, 从而

$$\lim_{m_{t1} \rightarrow \infty} \frac{1}{s_{m_{t1}}^3} \sum_{i=1}^{m_{t1}} E(|Y_{1i} - \mu_{1i}|^3) \geq 0.$$

综上所述, 存在

$$\lim_{m_{t1} \rightarrow \infty} \frac{1}{s_{m_{t1}}^3} \sum_{i=1}^{m_{t1}} E(|Y_{1i} - \mu_{1i}|^3) = 0.$$

即对 $k = 1$, 存在 ξ_1 ($\xi_1 = 1$) 使定理2中的 (11) 式成立, Y_{1i} 满足 Lyapunov 条件. 参考文献 [32] 可知, $\text{Sum}(S_{1t}) = \sum_{i=1}^{m_{t1}} Y_{1i}$ 满足中心极限定理, 即 $\text{Sum}(S_{1t})$ 服从正态分布.

当 k 取其他值时, 其分析方法同于第一层节点. 此处略去对其他层节点的分析, 直接跳至对第 ω 层节点的分析, 即 $k = \omega$.

第 ω 层节点 ($k = \omega$):

同于 $k = 1$, 对第 ω 层节点, 存在

$$u_{\omega i} = E(Y_{\omega i}) = q^\omega S_{\omega ti},$$

$$\sigma_{\omega i} = \text{Var}(Y_{\omega i}) = q^\omega(1 - q^\omega)S_{\omega ti}^\omega.$$

对 $\forall i (1 \leq i \leq m_{t\omega})$, 令 $\xi_\omega = 1$, 则有

$$\begin{aligned} & E\left(|Y_{\omega i} - u_{\omega i}|^{2+\xi}\right) \\ &= E\left(|Y_{\omega i} - u_{\omega i}|^3\right) \\ &= q^\omega(S_{\omega ti} - q^\omega S_{\omega ti})^3 + (1 - q^\omega)|0 - q^\omega S_{\omega ti}|^3 \\ &= q^\omega S_{\omega ti}^3(1 - q^\omega)(1 - 2q^\omega + 2q^{2\omega}), \end{aligned} \quad (15)$$

$$\begin{aligned} S_{m_{t\omega}}^{2+\xi} &= S_{m_{t\omega}}^3 \\ &= \sum_{i=1}^{m_{t\omega}} S_{\omega ti}^2 q^\omega(1 - q^\omega) \\ &\quad \times \sqrt{\sum_{i=1}^{m_{t\omega}} S_{\omega ti}^2 q^\omega(1 - q^\omega)}. \end{aligned} \quad (16)$$

结合 (15) 和 (16) 式, 可得

$$\begin{aligned} & \lim_{m_{t\omega} \rightarrow \infty} \frac{1}{S_{m_{t\omega}}^3} \sum_{i=1}^{m_{t\omega}} E\left(|Y_{\omega i} - u_{\omega i}|^3\right) \\ &= \frac{(1 - 2q^\omega + 2q^{2\omega})}{q^\omega(1 - q^\omega)} \lim_{m_{t\omega} \rightarrow \infty} \frac{\sum_{i=1}^{m_{t\omega}} S_{\omega ti}^3}{\sum_{i=1}^{m_{t\omega}} S_{\omega ti}^2 \sqrt{\sum_{i=1}^{m_{t\omega}} S_{\omega ti}^2}} \\ &\leq \frac{(1 - 2q^\omega + 2q^{2\omega})}{q^\omega(1 - q^\omega)} \frac{\text{sup}(S_{\omega t})^3}{\text{inf}(S_{\omega t})^3} \lim_{m_{t\omega} \rightarrow \infty} \frac{1}{\sqrt{m_{t\omega}}}. \end{aligned} \quad (17)$$

同理于 $k = 1$, 由 (17) 式可知, 对第 ω 层节点, 亦存在 $\xi_\omega (\xi_\omega = 1)$ 使定理 2 中的 (11) 式成立

$$\left(\lim_{m_{t\omega} \rightarrow \infty} \frac{1}{S_{m_{t\omega}}^3} \sum_{i=1}^{m_{t\omega}} E\left(|Y_{\omega i} - \mu_{\omega i}|^3\right) = 0 \right),$$

即 $Y_{\omega i}$ 满足 Lyapunov 条件, $\text{Sum}(S_{\omega t})$ 服从正态分布.

定理 2 表明, 网络中的 ω 组随机变量序列 $Y_{ki} (1 \leq K \leq \omega)$ 均满足 Lyapunov 条件, 即网络中各层节点的感知数据之和 $\text{Sum}(S_{kt}) = \sum_{i=1}^{m_{tk}} Y_{ki}$ 均服从正态分布. 又由于不同节点的感知数据能否成功传送至 Sink 节点是彼此独立互不影响的, 即 $\text{Sum}(S_t)$ 是 ω 个相互独立的正态分布变量 $\text{Sum}(S_{kt})$ 的和. 结合前文分析可知, $\text{Sum}(S_t)$ 将服从正态分布.

下面将探讨如何针对给定的相对误差值 ε 来计算 η 的值. 定理 3 通过将正态分布转化为标准正态分布后, 利用标准正态分布的相关特性来求解 η 的值.

定理 3 令 $\delta = 1 - \eta$, 设 $\varphi_{\delta/2}$ 是标准正态分布的 $\delta/2$ 分位数, 若 $\varphi_{\delta/2}$ 满足

$$\varphi_{\delta/2}^2 \leq \frac{q^\omega \text{inf}(N_t) \text{inf}(S_t) \varepsilon^2}{(1 - q^\omega) \text{sup}(S_t)}, \quad (18)$$

则此时网络得到的“近似聚集和” $\text{Sum}(S_t)$ 与“精确聚集和” $\text{Sum}(S_t)$ 之间的相对误差小于等于给定误差界限 ε 的概率将大于等于 η , 即

$$P\left(\left|\frac{\text{Sum}(S_t) - \text{Sum}(S_t)}{\text{Sum}(S_t)}\right| \leq \varepsilon\right) \geq \eta. \quad (19)$$

证明 由 (18) 式可得

$$\text{inf}(N_t) \text{inf}(S_t) \varepsilon^2 \geq \varphi_{\delta/2}^2 \text{sup}(S_t) \frac{(1 - q^\omega)}{q^\omega}.$$

又 $\text{inf}(N_t)$ 和 $\text{inf}(S_t)$ 分别表示 N_t 和网络中感知数据的最小值, 因此存在

$$\text{Sum}(S_t) = \sum_{i=1}^{N_t} s_{ti} \geq \text{inf}(N_t) \text{inf}(S_t).$$

所以

$$\varepsilon^2 \text{Sum}(S_t) \geq \varphi_{\delta/2}^2 \text{sup}(S_t) \frac{(1 - q^\omega)}{q^\omega}. \quad (20)$$

由定理 1 可知,

$$\text{Var}(\text{Sum}(S_t)) \leq \frac{(1 - q^\omega)}{q^\omega} \text{sup}(S_t) \text{Sum}(S_t),$$

$$E(\text{Sum}(S_t)) = \text{Sum}(S_t),$$

又 $\text{Sum}(S_t)$ 服从正态分布, 故由 (20) 可知

$$P\left(\left|\frac{\text{Sum}(S_t) - \text{Sum}(S_t)}{\text{Sum}(S_t)}\right| \leq \varepsilon\right) \geq \eta. \quad (21)$$

结合标准正态分布的中位数知识^[28]、(20) 式、(21) 式以及 $\delta = 1 - \eta$, 可得

$$P\left(\left|\frac{\text{Sum}(S_t) - \text{Sum}(S_t)}{\text{Sum}(S_t)}\right| \leq \varepsilon\right) \geq \eta. \quad (22)$$

5 R 算法的有效性验证

设一个传感器网络中包含 5000 个节点和 285 个 CH. 首先, 将上述 5000 个节点和 285 个 CH 随机布局在面积分别为 $5000 \text{ m} \times 5000 \text{ m}$, $10000 \text{ m} \times 10000 \text{ m}$, $15000 \text{ m} \times 15000 \text{ m}$ 的监测区域内. 设置网络中的 Sink 节点、CH 以及普通节点的初始能量为足够大 (即确保实验过程中不会出现节点死亡), 传播模型使用正态 shadowing 模型, 同时在簇首与簇首、簇首与 Sink 间的感知数据传输过程中加入超时重传机理, 网络模型中的其他参数设置同于文献 [33]. 在网络模型中运行 DMSTRP 协议, 将网络中的 Sink、CH 和传感器节点连成一个整体, 以仿真概率型 WSN 网络. 由于链路是以概率 (q) 来成功传送数据, 即便在活动节点数量、节点的布局、簇的结

构、簇的个数、单个簇内的普通节点数以及各相连节点间的距离(决定该链路的 q 值)等确定不变的情况下,网络在仿真时所得到的“近似聚集和”与“精确聚集和”间的相对误差亦存在不同.对仿真网络的“近似聚集和”与“精确聚集和”间的相对误差进行10000次试验,并针对不同的相对误差进行累计概率的统计,得到统计可信度,不同监测区域面积下的实验结果如图1至图3中的蓝色实线所示.

针对上述三个网络模型,通过统计各个簇的层数可得到各网络模型中簇层数的最大值 ω ;从网络中随机挑选出一个簇,并对该簇内相连节点间的距离进行统计,求得簇内相连节点间的平均距离,并用该距离来近似网络中相连节点间的距离 d ;将距离 d 代入(10)式中,得到该时刻下链路传送数据成功的概率 q .设置不同的相对误差界限 ε ,将 ε , q 和 ω 代入公式

$$\varphi_{\delta/2}^2 \leq \frac{q^\omega \inf(N_t) \inf(S_t) \varepsilon^2}{(1 - q^\omega) \sup(S_t)}$$

中可计算出各网络模型下的 η 值,称其为R算法可信度.对 η 与 ε 间的关系曲线进行描绘,得到如图1至图3中红色虚线.

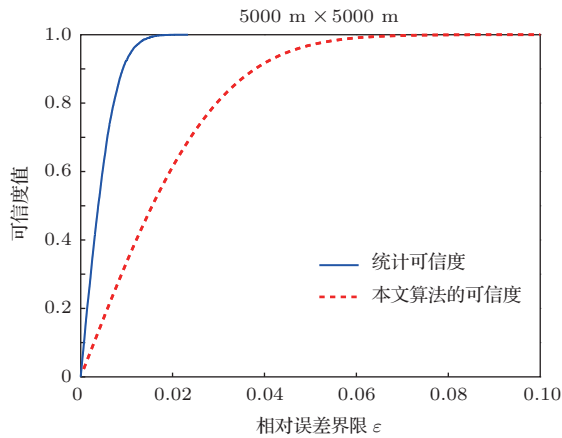


图1 (网刊彩色) R算法的有效性验证(5000 m × 5000 m)

考虑到可信度的最大值为1,仿真实验在描绘图1至图3中的曲线时,仅针对 η 位于0到1之间的实验结果进行了统计.

在监测区域面积设置为5000 m × 5000 m的网络模型中,网络仿真得到的统计可信度与R算法求得的R算法可信度如图1所示.对图1进行分析可得,统计可信度与R算法可信度的值均随 ε 的增大而增大.图中蓝色实线表明,统计可信度值在 ε 为0.0225时达到0.9999;而当R算法可信度达到0.9999时, ε 对应为0.0907.当监测区域面积为

10000 m × 10000 m时,统计可信度和R算法可信度的对比结果如图2所示.由图2中的蓝色实线和红色虚线可知,统计可信度值在 ε 为0.0367时达到0.9999,而R算法可信度达到0.9999时, ε 对应为0.1379.监测区域面积为15000 m × 15000 m时,实验结果如图3所示.图3的蓝色实线与红色虚线表明,当 ε 为0.0508时,统计可信度达到0.9999,而R算法可信度则是在 ε 为0.1768时达到0.9999.

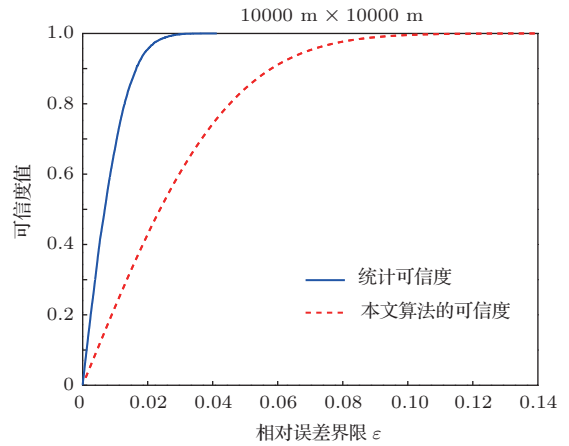


图2 (网刊彩色) R算法的有效性验证(10000 m × 10000 m)

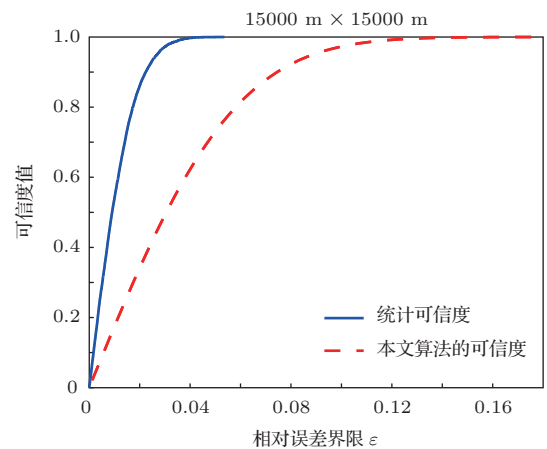


图3 (网刊彩色) R算法的有效性验证(15000 m × 15000 m)

上述数值结果表明,针对同一 ε 值,随着监测区域面积的增大,统计可信度与R算法可信度均呈现出降低趋势.这是由于,在网络节点总数和簇数一定的情况下,监测区域面积越大,网络中节点间的距离越远,链路中噪声对信号的影响程度随之增加;同时,节点间距离的增加使得网络中各节点的通信可达范围内的节点数逐渐减少,网络中簇的层数将增加,继而使得节点的感知数据被成功传送至Sink的平均概率减小,即Sink节点在一次数据

采集中所收到的感知数据将减少, 最终使得统计可信度降低. 图1至图3中的蓝色虚线还表明, 网络仿真所得到的“近似聚集和”与“精确聚集和”间的相对误差的最大值均较小, 即便在监测区域面积为 $15000\text{ m} \times 15000\text{ m}$ 的情况下, 最大相对误差值也仅为0.0507. 该现象表明, 本文所采用的数据聚集方法具有较好的近似效果, 同时也与前文证明“近似聚集和”是“精确聚集和”的无偏估计过程相对应. 此外, 由于R算法的目的在于求可信度的下限值, 在图1至图3中, 针对不同的相对误差界限 ε 均存在统计可信度值大于R算法可信度, 即验证了R算法的有效性. 然而, 由于在推导“近似聚集和”的方差上限时, 对方差范围进行的放缩处理, 使得R算法的精确程度不是很好.

6 实验结果的讨论

前一部分的实验, 其目的在于验证R算法的有效性, 接下来我们将对R算法的影响因素进行探讨.

在本节中, 第一组实验描述的是 η 随网络运行轮数的变化趋势. 本组实验是在第5节的网络模型基础上, 重设传感器节点的初始能量为2 J, 其他参数的设置不做改变. 同理于前文, 针对不同监测区域面积下的网络模型, 在网络运行过程中, 通过统计各个簇的层数得到模型中簇层数的最大值 ω ; 并用网络中某一簇内的相连节点间的平均距离来近似 d . 为方便观察 η 随网络运行轮数的变化, 设置三个网络模型中的 ε 分别为0.08, 0.14 和 0.25.

通过统计各个簇的层数可得到各网络模型中簇层数的最大值 ω ; 从网络中随机挑选出一个簇, 并对该簇内相连节点间的距离进行统计, 求得簇内相连节点间的平均距离, 并用该距离来近似网络中相连节点间的距离 d ; 将距离 d 代入(10)式中, 得到该时刻下链路传送数据成功的概率 q . 设置不同的相对误差界限 ε , 将不同运行轮数下的 q , ω 和 ε 代入公式

$$\varphi_{\delta/2}^2 \leq \frac{q^\omega \inf(N_t) \inf(S_t) \varepsilon^2}{(1 - q^\omega) \sup(S_t)}$$

中可得到各网络模型下的 η 值, 结果如图4, 图5和图6所示.

由于上述三个图中的曲线呈现出相同的变化趋势, 为此, 我们只针对图4中的曲线走势做具体分析.

图4描述的是监测区域面积为 $5000\text{ m} \times 5000\text{ m}$ 的仿真网络下 η 随网络运行轮数的变化趋势. 由于网络中簇的构建、感知数据的采集、数据在信道中的传输、节点对感知数据的转发、簇首节点对数据的超时重传以及簇首节点和Sink节点对感知数据的处理等均需要一定的时间, 使得在运行初期存在 $\eta = 0$ 的情况. 一旦Sink获得 q 和 ω 的值后 ($\omega = 3, q = 0.9308$), 将计算得到 η 的值, 从而使得 η 在第9轮处跳变为0.9611.

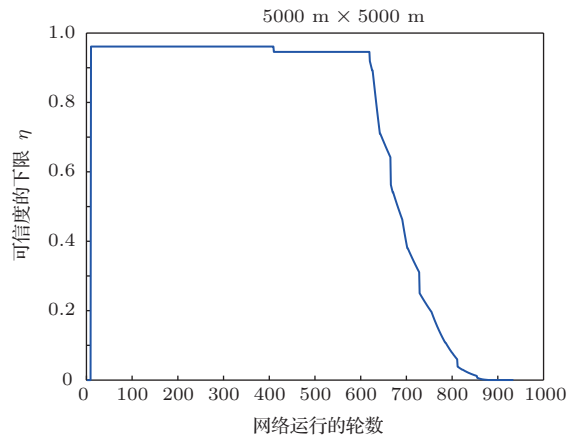


图4 η 的值随着网络运行轮数的变化趋势 ($5000\text{ m} \times 5000\text{ m}$)

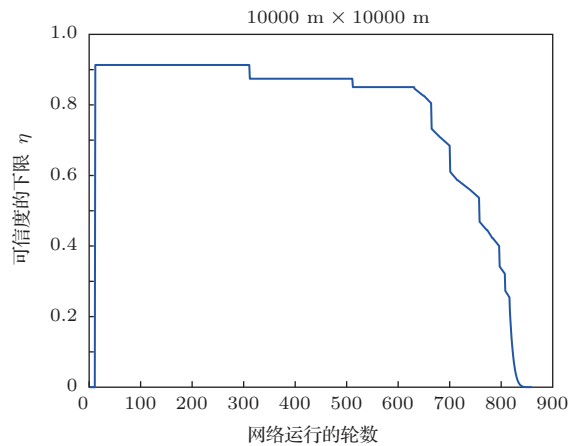


图5 η 的值随着网络运行轮数的变化趋势 ($10000\text{ m} \times 10000\text{ m}$)

前文的分析表明, 针对某一网络轮数, q 和 ω 的值主要取决于当前的网络规模、活动节点数量、簇的内部结构等. 由于上述三个图描述的是同一固定网络规模下的仿真结果, 即对单个网络模型而言, 不存在网络规模对 q 和 ω 值的影响. 由于模型中所散布的簇首节点的初始能量远大于传感器节点, 依据文献[27]可知, DMSTRP协议在运行过程中, 选定的簇首节点均为初始布置的CH; 而剩余能量较

多的传感器节点将在簇的内部充当父节点的角色;随着网络运行轮数的增加,不同的能耗使得各传感器节点的剩余能量发生变化,从而使得网络结构在簇的重构过程中发生变化.通过分析图4的仿真结果可得,在初始的610轮中,网络中的活动节点数均为5000.由于网络在靠近400轮处发生了一次簇的重构,使得网络中相连节点间的链路距离以及簇的层数均发生改变,链路距离的改变使得链路中噪声对信号的影响程度发生变化,同时,簇层数的不同使得网络中节点的感知数据被成功传送至Sink节点的平均概率亦被改变,从而使得 η 在第409轮处跳变为0.9457,并持续该数值长达210轮.其次,由于网络中活动节点数目在600轮处开始减少,且DMSTRP协议中簇的建立阶段和稳定阶段在网络运行期间是交替循环的,固定区域内节点数的减少将使得节点间的距离随即增大;在发射功率不变的情况下,各节点通信可达范围内的节点数将减少,簇内节点的层数继而增加;即 q 和 ω 的值将随网络中节点数的减少而变化,最终导致 η 从619轮开始呈现出下降趋势直至0.此外,我们还观察到,在后续811轮至934轮期间, η 的值都较低,这是由于在网络运行的后续周期中,活动节点数目都保持在一个较低的水平;当较少的活动节点布局于规模一定的网络中时,模型所对应的 ω 值和节点间距离将增大,而节点间距离的增大使得 q 值变小,从而使得R算法计算出的 η 值较低.

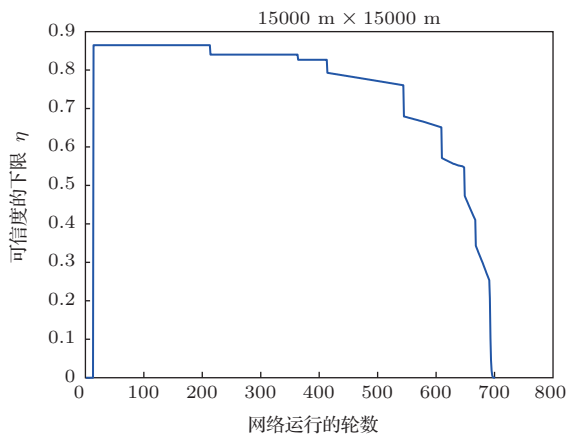


图6 η 的值随着网络运行轮数的变化趋势 (15000 m × 15000 m)

在第一组实验的基础上,我们接下来将针对三个网络模型的延迟进行分析.第一组实验结果表明,随着运行轮数的增加,网络模型中的活动节点数在逐渐减少,当活动节点数减少至0时,网络运行结束.为此,在本组实验中,将描述网络延迟随

着活动节点数的变化趋势.

此处,定义网络延迟为,从Sink节点发出聚集命令开始,到所有感知数据汇聚到Sink节点所花费的时间.延时的单位(delay)设置为数据包在两个节点间链路上的传送时间^[34].以图7中简单的簇为例,当4号节点发出聚集命令后,在第一周期中,1和8号节点会将其感知数据传送至其父节点,第二周期中,2号和7号将其自身的感知数据以及所收到的来自其他节点的数据传送至其父节点,第三周期为3号和6号节点传送数据,第4周期为5号节点传送数据.即在图7中,网络延迟对应为4 delay.当簇首依据超时重传的方式来保证感知数据的正确传送时,一次超时重传,将使得网络延迟增加1个delay.

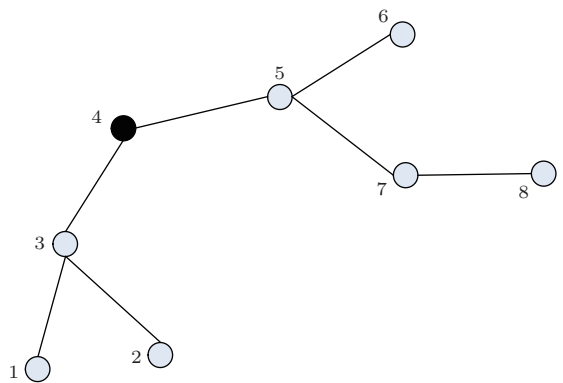


图7 一个简单的簇

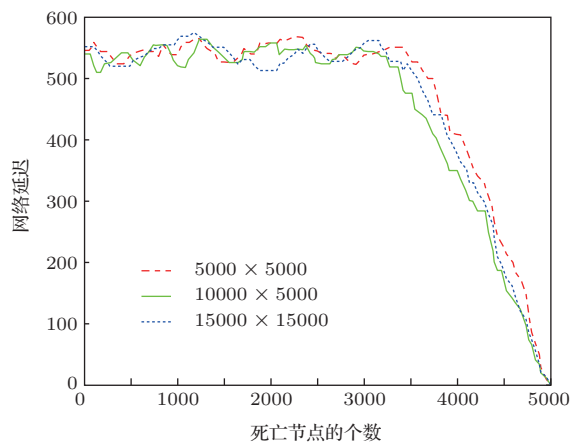


图8 不同规模下的网络延迟

针对第一组实验中的三个网络模型,对其运行过程中的网络延迟进行统计,将延迟结果统计到一个图中,得到如图8所示的结果.由于网络在运行过程中,传感器节点剩余能量的变化、网络中簇的

重构、簇首节点对感知数据的超时重传等因素,使得网络延迟呈现出如图 8 所示的变化趋势. 图 8 表明,当网络中的活动节点数在 1500 以上时,网络延迟在 524 上下波动,而当网络中的活动节点数低于 1500 时,网络延迟将呈现出减小趋势,直至为 0. 这是由于,当网络中存在较多的活动节点时,网络中的链路数较多,簇的个数以及簇中节点的层数均较多,使得一次数据聚集所需的周期数较多,即延迟较大.

最后一组实验,我们将前文的 5000 个传感器节点和 285 个 CH 布局在面积不同的监测区域内,以分析网络规模对 R 算法的影响. 其中监测区域面积的变化区间为 5000 m × 5000 m 到 15000 m × 15000 m,并在其中取出 15 个面积值作为实验的测试对象. 公式

$$\varphi_{\delta/2}^2 \leq \frac{q^\omega \inf(N_t) \inf(S_t) \varepsilon^2}{(1 - q^\omega) \sup(S_t)}$$

表明, η 值的求解需要给定 ε , q 和 ω 的值. 又当网络所运行的路由协议一定时, q 和 ω 的值将主要取决于网络规模,且 ε 是依据网络用户的应用需求而定,为此, η 将随着网络规模的变化而变化. 当 ε 分别设置为 0.01, 0.03 和 0.05 时, η 的变化曲线如图 9 所示. 随着网络面积的增大,网络中节点间的链路距离随即增加,链路中噪声对信号的影响程度继而增大;此外,不同网络面积下,传感器节点的通信可达范围内的节点数不同,使得各簇的结构亦发生改变,簇内节点层数的不同使得网络中节点的感知数据被成功传送至 Sink 的平均概率发生变化;最终使得 R 算法的计算结果随网络面积呈现出了图 9 的变化趋势.

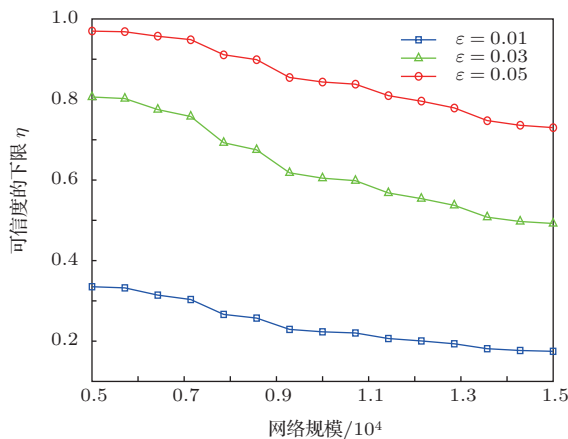


图 9 网络规模对 R 算法的影响

7 结 论

由于大规模 WSN 环境监测区域内用于传输和处理数据的物理资源受限,以及过渡区的信噪比影响,使得有损链路下的数据采集与处理呈现出一种饱和流状态,网络无法向用户提供精确的查询结果,最终导致网络用户不能对网络中的感知数据进行充分的分析和利用. 考虑到 WSNs 中节点的能量受限,本文采用动态最小生成树路由协议来构建基于簇的网络模型;并对过渡区状态下的 WSN,选用“聚集和”操作符,通过对网络中的感知数据进行“聚集和”操作,得到“近似聚集和”Sum(S_t). 为指导用户对接收到的 Sum(S_t) 做深度处理(是将 Sum(S_t) 作为决策依据,又或参考依据,还是直接丢弃此次的聚集结果并再次聚集)和提高网络的感知性能,本文引入可信度 η 参数,用“近似聚集和”与“精确聚集和”间的相对误差小于等于某一误差界限的最小概率来衡量近似聚集结果的可信度,并将网络的数据聚集结果和 η 参数一同传送至用户.

仿真实验验证了 R 算法的有效性,对比 R 算法的计算结果 η 曲线和仿真实验所统计出的可信度曲线可知,针对任意相对误差界限 ε 下, R 算法所得到的可信度下限 η 均小于仿真实验的统计可信度值;描述了过渡区内由于信噪比导致的链路不可靠所引起的可信度 η 的变化规律,信噪比越低,链路传送数据成功的概率 q 越小,网络中近似聚集和的可信度 η 将越低;讨论了网络性能对可信度 η 的影响,针对某一具体网络,随着网络运行周期的增加,可信度 η 由 > 0.9 逐渐降低到 0;分析了同一相对误差界限 ε 下,可信度 η 随着网络规模的变化趋势,网络规模越大,可信度 η 越低;并对不同规模下的网络延迟进行统计;从而为大规模 WSNs 从理论模型投入实际工业现场应用提供了理论依据和经验公式.

参考文献

- [1] Pei H, Li X, Mutka M W, Ning X 2013 *IEEE Communications Surveys & Tutorials* **15** 101
- [2] Rozyyev A, Hasbullah H, Subhan F 2011 *Research Journal of Information Technology* **3** 81
- [3] Szweczyk R, Osterweil E, Polastre J, Hamilton M, Mainwaring A, Estrin D 2004 *Communications of the ACM* **47** 34
- [4] Chauhdary S H, Bashir A K, Shah S C, Park M S 2009 *Journal of Applied Sciences* **9** 4247

- [5] Biswas P K, Phoha S 2006 *IEEE Transactions on Computers* **55** 1033
- [6] Lee L T, Chen C W 2008 *Information Technology Journal* **7** 737
- [7] Sabri N, Aljunid S A, Ahmad B, Yahya A, Kamaruddin R, Salim M S 2011 *Journal of Applied Sciences* **11** 3104
- [8] Kumar D 2011 *Research Journal of Environmental Sciences* **5** 105
- [9] Yick J, Mukherjee B, Ghosal D 2008 *Computer Networks* **52** 2292
- [10] Arampatzis T, Lygeros J, Manesis S 2005 *Proceedings of the 20th IEEE International Symposium on Intelligent Control (ISIC '05)* 719
- [11] Tseng Y C, Pan M S, Tsai Y Y 2006 *Computer* **39** 55
- [12] J Wu, S F Yuan, Zhao X, Yin Y, Ye W S 2007 *Chin. Phys. B* **16** 1898
- [13] Hao Y, Foster R 2008 *Chin. Phys. B* **29** R27
- [14] Jang S D, Kang B W, Kim J 2013 *Chin. Phys. B* **22** 025002
- [15] Wang C L, De D, Song W Z 2013 *Knowledge-based Systems* **37** 346
- [16] Deligiannakis A, Kotidis Y, Rossopoulos N 2006 *Information Systems* **31** 770
- [17] Guo J L 2010 *Chin. Phys. B* **19** 120503
- [18] Tan R, Xing G L, Yuan Z H, Liu X, Yao J G 2013 *ACM Transactions on Sensor Networks* **9** 28:1
- [19] Cong X, Liu S L, Ma R 2008 *Acta Phys. Sin.* **57** 7487 (in Chinese)[丛蕊, 刘树林, 马锐 2008 物理学报 **57** 7487]
- [20] Wang P, He Y, L, Huang S 2013 *Ad hoc Networks* **11** 1287
- [21] Lu Z Q, Tan S L, Biswas J 2013 *Wireless Personal Communications* **70** 391
- [22] Wang Y Q, Yang X Y 2012 *Acta Phys. Sin.* **61** 090202 (in Chinese) [王亚奇, 杨晓云 2012 物理学报 **61** 090202]
- [23] Liu Y, Zhang Q, Ni L M 2010 *IEEE Transactions on Parallel and Distributed Systems* **21** 405
- [24] Wan P J, Huang S C, Wang L, Wan Z, Jia X 2009 *The Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing* 185
- [25] Akyildiz I F, Su W, Sankarasubramaniam Y, Cyirci E 2002 *Computer Networks* **38** 393
- [26] Huang G Y, Li X W, He J 2006 *IEEE Conference on Industrial Electronics and Applications, Singapore* 1
- [27] Muruganathan S D, Ma C F, Bhasin R I, Fapojuwo A O 2005 *IEEE Communications Magazine* **43** 8
- [28] Shen Z, Xie S Q, Pan C Y 2005 *Probability Theory & Mathematical Statistics* (Beijing: Higher Education Press) (in Chinese) [盛骤, 谢式千, 潘承毅 2005 概率论与数理统计 (北京: 高等教育出版社)]
- [29] Bernstein S, Bernstein R 2004 *1st ed. Columbus: McGraw-Hill*
- [30] Fazio P, De R F, Sottile C 2011 *Performance Evaluation of Computer & Telecommunication Systems (SPECTS)* 98
- [31] Zuniga M, Krishnamachari B 2004 *IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks* 517
- [32] Fischer H 2011 *1st ed. New York Springer* 194
- [33] He S M 2013 *Ph. D. Dissertation* (Changsha: Hunan University) [何施茗 2013 博士学位论文 (长沙: 湖南大学)]
- [34] Lindsey S, Raghavendra C, Sivalingam K M 2002 *IEEE Transactions on Parallel and Distributed Systems* **13** 924

Data fusaggregation algorithm based on dynamic minimal spanning tree routing protocol*

Peng Hai-Xia[†] Zhao Hai Li Da-Zhou Lin Chuan

(Institute of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

(Received 13 December 2013; revised manuscript received 20 January 2014)

Abstract

Large-scale geographic environmental monitoring and physical infrastructure for transferring and processing data cannot maintain the same growth rate with the monitoring scale, thus make the data gathering and processing under unreliable links, showing a kind of saturation flow state; and the ability of wireless sensor network (WSN) seemingly cannot be stabilized. On the other hand, though the calculation results from ideal network model are accurate enough, they deviate from the practical application greatly, hence the network users cannot adequately analyze and utilize the sensed data from industrial field network, and also cannot analyze the influence of network size and performance on data fusaggregation. Because of these, we will present the ‘transitional region phenomenon’ to be one of assumptions for industrial field simulation, and propose a data fusaggregation algorithm for the practical application on this basis, i.e. data fusaggregation algorithm based on reliability (R algorithm). When designing the R algorithm, as an example of sum operator, the lower limit η of reliability of appreciate aggregation sum result will be calculated by analyzing and synthesizing the result automatically. Then the aggregation sum result and the value of η will be sent to users together. In addition to providing the summary information from the monitored area to users, R algorithm also provides the parameter η as the judgment of information reliability to facilitate users to do further handling of aggregation results and improve the WSN sensing performance. Simulation results describe the changing rule of reliability η caused by unreliable links from the signal-to-noise ratio in transitional region, and discuss the network influence of size and performance on reliability η , with the increase of network operation cycles and network scale, when the value of reliability η becomes gradually close to 0. And this provides theoretical foundations and empirical formulas for WSNs from theoretical model to practical industrial application.

Keywords: wireless sensor networks, data fusaggregation, DMSTRP, reliability

PACS: 02.50.-r, 02.50.Sk, 05.60.-k, 06.20.Dk

DOI: 10.7498/aps.63.090206

* Project supported by the funding from the National Science and Technology Support Program (Grant No. 2012BAH82F04), the National Natural Science Foundation of China (Grant No. 61101121), and the National High Technology Research and Development Program of China (Grant No. 2013AA102505).

[†] Corresponding author. E-mail: penghaixia@neueru.com