

## 无线传感器网络事件簇的数据聚集容错机制

肖伟<sup>1,2</sup>, 徐明<sup>1</sup>, 吕品<sup>1</sup>, 余建平<sup>2</sup>

(1. 国防科技大学 计算机学院, 湖南 长沙 410073; 2. 湖南师范大学 数学与计算机学院, 湖南 长沙 410081)

**摘 要:** 针对数据聚集操作的容错要求, 提出了事件簇的数据聚集容错机制 EFSA。在生成事件簇的基础上, 采用  $k$ -means 算法提取加权平均数作为近似的事件值, 并且计算和迭代地调整节点的可信度, 作为聚集计算的数据权值和节点是否出现数据错误的指标。分析和实验表明, EFSA 机制能够削弱数据错误的影响, 实现数据聚集的正确性与可靠性。

**关键词:** 无线传感器网络; 事件簇; 容错; 数据聚集;  $k$ -means 算法

中图分类号: TP393

文献标识码: B

文章编号: 1000-436X(2010)06-0112-07

## Fault tolerant scheme for data aggregation in event cluster over wireless sensor networks

XIAO Wei<sup>1,2</sup>, XU Ming<sup>1</sup>, LV Pin<sup>1</sup>, YU Jian-ping<sup>2</sup>

(1. School of Computer, National University of Defense Technology, Changsha 410073, China;

2. College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China)

**Abstract:** A fault tolerant scheme for data aggregation based event clustering called EFSA was proposed, based on fault tolerant capability in data aggregation over wireless sensor networks. The event cluster was generated before the weighted average data were extracted as approximate event value by way of  $k$ -means algorithm. The confident rate of every node in the event cluster was computed and adjusted by iterative algorithm and the CR functioned as the weighted value in aggregating data and the indicator of data fault from nodes. Simulation results and analysis demonstrate that EFSA enhances correctness and reliability of data aggregation by weakening the influence from data failure.

**Key words:** wireless sensor networks; event clustering; fault tolerant; data aggregation;  $k$ -means algorithm

### 1 引言

事件监测是无线传感器网络最广泛的应用之一<sup>[1,2]</sup>。当有事件发生时, 能感知该事件的传感器节点构成事件簇<sup>[3]</sup>, 事件簇的簇头节点对传感器节点收集的数据进行网内聚集操作<sup>[4-6]</sup>, 然后再传递给基站节点或者最终用户。但是由于节点易错性及能量资源限制, 设计一种高效、容错的数据聚集策略对传感器网络非常重要。

针对数据聚集操作的容错问题, 国内外机构给出了一些相关的研究进展: Desovski D<sup>[7]</sup>等给出的算法在错误发生指数分布的假设条件下, 使用马尔科夫模型计算系统和随机投票的方式进行数据聚集操作; Nathy S<sup>[8]</sup>等给出了聚集架构 synopsis diffusion, 它使用能量高效多路径路由结构来进行数据收集及容错处理; Gobriel S<sup>[9]</sup>等研究了 2 种基于复件—感知聚集的容错模式, 利用冗余路径来传送正确的聚集结果; Wang Y<sup>[10]</sup>等提出了一种延迟/容错数据聚集结

收稿日期: 2009-11-15; 修回日期: 2010-04-28

基金项目: 国家自然科学基金资助项目 (60773017, 60903168)

**Foundation Item:** The National Natural Science Foundation of China(60773017, 60903168)

构, 利用参数传输率来决定传输数据分组的时机, 利用容错率确定数据分组复制的次数; Kempe D<sup>[11]</sup>研究了 Tree 数据聚集的局限性, 提出 Gossip 的容错数据聚集协议, 要求每个节点存储权值和融合值, 在每一轮操作过程中, 随机选择一个邻居传送它的一半权值和一半融合值, 在特定的网络状况下得到收敛的融合值; Chen J Y<sup>[12]</sup>等给出了分布式随机分组协议, 它随机选择广播节点集合, 然后计算出局部聚集的值。

不同于以上的研究, 本文采用随机产生的事件簇结构, 针对节点出现的数据错误, 在数据聚集操作中提出了一种事件簇的数据聚集容错机制 EFSA(event-clustering based fault-tolerant scheme for data aggregation)。EFSA 在分布生成事件簇的基础上, 簇头节点基于  $k$ -means 算法, 利用节点可信度 CR(confident rate)获取近似事件值, 减少了错误数据对数据聚集的影响。此外, EFSA 通过对事件簇内节点可信度的迭代调整, 达到出错节点的可信度迅速降低的目的, 结果既保证了带权数据聚集计算的正确性, 又能够给出节点数据是否出错的判断。

本文第2节给出了 EFSA 主要组成及实现过程; 第3节给出 EFSA 性能的理论分析及参数讨论; 第4节给出仿真实验及其讨论; 第5节是本文的结束语。

## 2 EFSA 事件簇数据聚集容错机制

### 2.1 系统所用模型

假设一个高密度、均匀分布的无线传感器网络  $G(V, E)$ ,  $V$  为节点集合,  $E$  为边的集合。每个节点都有唯一的 ID 号, 节点的工作状态为睡眠状态或者激活状态, 没有事件感知的状态为睡眠状态。无线传感器网络中事件簇的数据处理模型如图1所示, 当有事件发生时, 事件感知区域是以事件(event)为中心, 半径为感知半径的一个圆形区域, 区域内的传感器节点构成一个事件簇。

在本文中, 传感器和 sink 节点的相互作用采用的是发布/订阅<sup>[13]</sup>模型, 能量消耗采用 LEACH<sup>[14]</sup>模型, 具体的能量消耗公式参考文献[14]。主要错误为节点数据错误, 包括: ①节点感知错误, 节点感知了事件但报告的事件数据误差较大; ②节点漏报错误, 在指定的时间内, 节点感知事件但没有报告数据给 sink 节点; ③节点误报错误, 报告了一个在它感知半径以外的事件, 或者感知半径内没有发生的事件。假设每个节点的出错概率  $p$  相同, 节点出错相互不干扰。

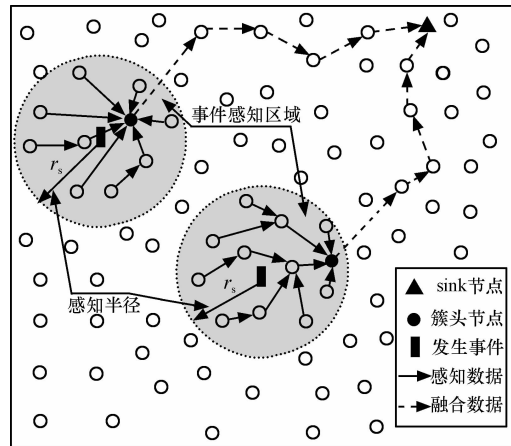


图1 事件簇数据处理模型

### 2.2 事件簇的分布式构造

类似于文献[15]给出的可信指标概念, 引入可信度 CR 作为节点数据在数据聚集操作中的可信程度和衡量节点是否出现数据错误的指标。可信度 CR 定义为 0 到 1 之间的实数, 初值为 1, 每个节点的可信度通过簇头节点计算并动态赋予各节点。

如图1所示, 当传感器网络中有某个事件发生时, 感知区域内的节点被激活, 并通过选举算法构成事件簇。本文中, 采用类似于 ARPEES<sup>[3]</sup>的算法, 利用事件簇中节点的剩余能量与可信度的结合来选举簇头, 目的在于既保证网络具有最大生命周期, 又能避免错误数据对数据聚集的影响。

初始过程中, 每个节点保持数据结构  $(E_{res}(i), CR(i), D(i))$ , 分别表示节点  $i$  剩余能量、节点  $i$  可信度和节点  $i$  的感知事件值。当感知的数据信号强度超过预先设定数据阈值时(不同的事件对应的数据阈值不同), 节点就变成了激活状态, 利用文献[3]提供的方法, 激活的节点广播 REQ-CLUS 信息给它们的邻居节点。这个信息中包括了节点的 ID 号, 剩余能量的值和节点可信度。即

$$REQ\_CLUS(ID(i), E_{res}(i), CR(i)) \quad (1)$$

信息发送以后, 每个节点在限定时间  $t$  内接收到其他激活节点发送过来的 REQ-CLUS 信息, 然后按式(2)计算每个节点的  $F_{CH}(i)$ ,  $X$  表示节点集合。

$$F_{CH}(i) = E_{res}(i) \times CR(i), \quad i \in X \quad (2)$$

具有最大  $F_{CH}(i)$  值的节点设置自己作为簇头, 然后通知所有的激活节点, 构造了一个事件簇。当簇头能量低于某个阈值, 或者簇头节点出现了错误时, 事件簇的簇头节点将按以上的选举方法重新选

举簇头。很明显，初始状态时，所有节点的可信度为 1，上述的簇生成算法就转化为 ARPEES 算法。

### 2.3 利用 $k$ -means 算法计算事件的近似值

为了削减错误数据对事件值的影响，采用  $k$ -means<sup>[16]</sup>方法对数据进行分割集合操作，并利用节点具有的可信度值确定每个数据集合可信度，最终选择具有最大集合可信度的数据集合来计算事件近似感知值。

假设事件簇中节点数为  $n$ ，获得的数据分别为  $D=\{d_1, d_2, \dots, d_n\}$ ，相应的可信度集合为  $R=\{CR(1), CR(2), \dots, CR(n)\}$ ， $\text{dist}(x, y)$  表示 2 个数据之间的距离， $E\_tole$  设为数据聚集集合半径作为分割的参数。 $k$ -means 算法的目标是将  $D$  中的数据分别放入独立不相交的集合  $C=\{C_1, C_2, \dots, C_r\}$  中，其中  $C_i$  是数据集合，并且满足条件  $\cup C_i = D$  且  $C_i \cap C_j = \emptyset$ ， $\forall i, j \in 1, \dots, r$ 。同时设  $\{c_1, c_2, \dots, c_r\}$  为  $C$  的算术平均值集合。构造集合  $C$  的  $k$ -means 算法如下。

**步骤 1** 初始化过程，设  $C$  为空集，首先将数据集  $D$  内数据排成一个升序队列，取  $D$  内的最小和最大的数据  $d_1$  和  $d_2$  分别放入集合  $C_1$  和  $C_2$  中，将  $C_1$  和  $C_2$  加入  $C$  并计算出  $c_1$  和  $c_2$ 。

**步骤 2** 对任意数据  $d_k$ ，如果满足  $\text{dist}(d_k, c_i) > E\_tole$ ， $\forall C_i \in C$ ，则创建一个新的集合  $C_k$  加入  $C$ ，而且  $d_k \in C_k$  并且计算出  $c_k$ 。

**步骤 3** 如果存在多个数据集合与  $d_k$  的距离小于  $E\_tole$ ，则选择与  $d_k$  最近的集合  $C_m$ ，使  $d_k$  加入集合  $C_m$ 。

**步骤 4** 重复对数据集  $D$  执行步骤 2 和步骤 3，使所有的数据  $d \in D$  都加入某个集合  $C_i$ ，并且得到集合  $C$ 。如果 2 个或者多个集合  $C_i \in C$  中值的相互距离小于  $E\_tole$ ，则重新合并这些集合为一个集合，并重新计算中值。重复执行这个过程，直到集合  $C$  不再发生变化为止。

对于获得的集合  $C=\{C_1, C_2, \dots, C_r\}$ ，计算每个集合  $C_i$  中节点的可信度平均值  $ACR_i$ ，如果数据集合  $k$  对应的  $ACR_k$  值最大，利用式(3)求出集合  $C_k$  的加权平均数作为事件近似感知数据值  $Eventavg$ ，设  $M$  为集合  $C_k$  元素个数， $C_k = \{d_{k1}, d_{k2}, \dots, d_{kM}\}$ 。

$$Eventavg = \frac{\sum_{j=1}^M (CR(kj))}{\sum_{i=1}^M CR(ki)} d_{kj} \quad (3)$$

由文献[16]的研究可以得出，以上的  $k$ -means 算法的时间复杂度为  $O(n)$ 。以下章节将对参数

$E\_tole$  的设定进行讨论。

### 2.4 簇内节点可信度的调整及错误节点的确定

在获得事件近似值  $Eventavg$  以后，簇头节点将对每个节点的数据与  $Eventavg$  偏离程度进行计算，由此获取节点数据的可信度，作为节点新的可信权值。本文结合文献[15]提供的方法来计算节点的可信度。在计算过程中，簇头节点为事件簇的每个成员准备和维持一个变量  $y_i$ ，初始值都为 0。如果簇头节点发生了转移，变量  $y_i$  也同时传递给新的簇头。

首先计算每个节点  $i$  的获得值与平均值之间的距离  $m_i$ ，然后利用  $m_i$  与数据集合聚集半径  $E\_tole$  的关系及计算函数(式(4))得到当前  $y_i$  值。

$$y_{i,curr} = \begin{cases} 0, & m_i \leq \epsilon_{tole}, y_{i,old} - p \leq 0 \\ y_{i,old} - p, & m_i \leq \epsilon_{tole}, y_{i,old} - p > 0 \\ y_{i,old} + 1 - p, & m_i > \epsilon_{tole} \end{cases} \quad (4)$$

其中， $y_{i,old}$  为节点  $y_i$  的原数值， $y_{i,curr}$  为节点  $y_i$  的当前值， $p$  为节点出错概率。

最后，利用式(5)求出每个节点的可信度的值。

$$CR(i) = a^{-\lambda y_i} \quad (5)$$

其中， $a$ 、 $\lambda$  是比例常量，依赖于实验及应用，下面讨论 2 个参数的取值。

为了能够判定出错的节点，设定错误阈值  $min\_error$ ，如果节点的可信度  $CR(i)$  小于  $min\_error$ ，可以认为该节点为错误节点，然后把错误节点的 ID 号告知事件簇的各节点，完成错误节点的确定及隔离。如果簇头节点出现错误，发出新簇头节点的选举操作消息，按 2.2 节提供的选举算法构造新的事件簇。

## 3 EFSA 性能的理论分析及参数讨论

在给定的无线传感器网络中，假设有相对独立的  $n$  个节点构成事件簇，每个节点  $i$  的取值  $x_i$ ，满足分布  $N(A, \sigma^2)$  (其中  $\sigma$  取值非常小)，每个节点  $i$  对应的可信度  $r_i$  作为权值，初始值为 1。 $n$  个节点对感知的事件周期性地取样，每次取样以后，簇头节点对  $n$  个节点的感知数值进行迭代聚集操作，并且每个节点的出错概率相同且独立，节点出现数据错误的概率比较小，属于小概率事件。

### 3.1 EFSA 的概率收敛性能

首先讨论簇头节点不采用  $k$ -means 方法，对所有节点数据进行含权均值计算的收敛性能。假设第  $k$  次迭代时，采样的样本值为  $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}$ ，通过

式(4)、式(5)求得每个节点的可信度  $r_1^{(k)}, r_2^{(k)}, \dots, r_n^{(k)}$ ，其中  $0 < r_i^{(k)} \leq 1$ ， $i=1, 2, \dots, n$ ， $r_i^{(k)}$  体现了  $x_i^{(k)}$  与  $\tilde{X}^{(k-1)}$  的偏离程度，令  $gr_j^{(k)} = r_j^{(k)} / \sum_{i=1}^n r_i^{(k)}$  满足  $\sum_{j=1}^n gr_j^{(k)} = 1$ ，计算均值采用的公式为

$$\tilde{X}^{(k)} = \sum_{j=1}^n gr_j^{(k)} x_j^{(k)} \quad (6)$$

**引理 1** 在 EFSA 机制中，第  $k$  次迭代获得的均值方差不大于初始迭代的均值方差。

**证明** 由迭代过程可以有：

$$\begin{aligned} D(\tilde{X}^{(1)}) &= D\left(\sum_{j=1}^n x_j^{(1)} / n\right) \\ &= \frac{1}{n^2} \sum_{j=1}^n [D(x_j^{(1)})] = \frac{1}{n^2} n \sigma^2 = \frac{1}{n} \sigma^2 \quad (7) \end{aligned}$$

$$D(\tilde{X}^{(k)}) = D\left(\sum_{j=1}^n gr_j^{(k)} x_j^{(k)}\right) = \left(\sum_{j=1}^n (gr_j^{(k)})^2\right) \sigma^2 \quad (8)$$

利用不等式：

$$\begin{aligned} (a_1^2 + a_2^2 + \dots + a_n^2) &\leq \frac{(a_1 + a_2 + \dots + a_n)^2}{n}, \\ a_i > 0, i &= 1, 2, \dots, n \end{aligned}$$

$$\begin{aligned} D(\tilde{X}^{(k)}) &= \left(\sum_{j=1}^n (gr_j^{(k)})^2\right) \sigma^2 \\ &\leq \left(\frac{(gr_1^{(k)} + gr_2^{(k)} + \dots + gr_n^{(k)})^2}{n}\right) \sigma^2 = \frac{1}{n} \sigma^2 = D(\tilde{X}^{(1)}) \quad (9) \end{aligned}$$

即  $D(\tilde{X}^{(k)}) \leq D(\tilde{X}^{(1)})$ ， $k=2, 3, \dots, n$  □

**定理 1** 事件簇中根据 EFSA 机制，获得的事件均值随着迭代次数的增加更接近于期望值  $A$ ，并且当迭代次数趋向于无穷时，事件均值依概率收敛于  $A$ 。

**证明** 因为  $x_i^{(k)} \sim N(A, \sigma^2)$ ，同时，由引理 1 可以得出，对于每次迭代过程中的样本方差满足： $\bar{S}^{(k)} \leq \bar{S}^{(k-1)} \leq \dots \leq \bar{S}^{(1)}$ ， $\bar{S}^{(k)}$  表示第  $k$  次的样本方差，即  $\bar{S}^{(k)} = \frac{1}{n} [(X^{(k)} - x_1^{(k)})^2 + \dots + (X^{(k)} - x_n^{(k)})^2]$ 。

同时从 EFSA 迭代过程可知，正确节点的权值保持较高的可信度值，而出错节点的权值会越来越小。正确节点的数据对  $\tilde{X}^{(k)}$  的影响程度会越来越大，而出错节点数据对  $\tilde{X}^{(k)}$  的影响越来越小，所以

$X^{(k)}$  会更接近于期望值  $A$ 。即对于  $\forall \epsilon$ ，都有  $|\tilde{X}^{(k)} - A| \leq |\tilde{X}^{(k-1)} - A| \leq \dots \leq \epsilon$ 。

当  $k \rightarrow \infty$  时，由小概率事件原理可知，出错节点的概率为零。也就是说，出错节点的权值越来越小，并趋近于零。因此可视出错节点在无穷次迭代后几乎不发生，即

$$\lim_{k \rightarrow \infty} \tilde{X}^{(k)} \xrightarrow{P} A \quad \square$$

在 EFSA 机制中，对于第  $k$  次迭代过程，如果采用  $k$ -means 方法时，从  $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}$  中选取的子集具有最大平均可信度权值，按  $k$ -means 方法和小概率事件原理，获得的均值  $\tilde{X}^{(k)*}$  比  $\tilde{X}^{(k)}$  更接近于期望值  $A$ ，同样满足以概率收敛于  $A$ ，即  $\lim_{k \rightarrow \infty} \tilde{X}^{(k)*} \xrightarrow{P} A$ 。

### 3.2 EFSA 参数的讨论及选择

从 2.3 节分析可以得出， $E\_tole$  的选择对  $k$ -means 中集合的产生个数有重要的作用，同时也影响最后的分割结果。固定节点出错比例，讨论在不同的数据集合半径  $E\_tole$  下，不同节点数的事件簇获得的事件值与真实值之间的相对误差变化。

如图 2 所示，聚集数据的相对误差开始时随  $E\_tole$  的增加逐渐减少，到了一定区间以后相对误差反而加速上升。原因在于  $E\_tole$  开始增大时，参加数据聚集操作的节点数增多，正常节点数的增多提高了聚集操作数据的正确性。然而随着  $E\_tole$  的继续增大，数据聚集操作中出错节点参加的数量也增加很快，造成了聚集数据的正确性反而下降。所以选择合适的  $E\_tole$  对最终事件值的正确获取有关键的作用。实验表明， $E\_tole$  选择 0.6~1.0 之间的值时，可以获得比较小的相对误差，同时也能尽可能减少错误节点对数据聚集操作的影响。

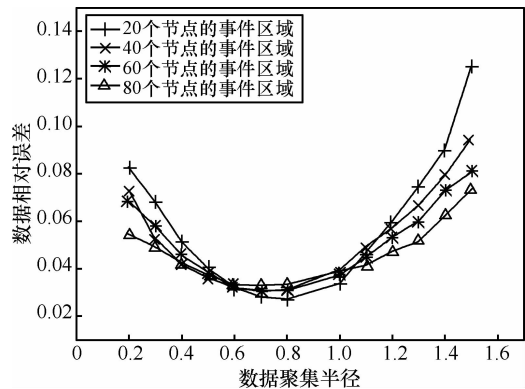


图 2 不同节点数对应的相对误差比较

接着考虑可信度计算公式  $CR = a^{-\lambda x}$  的参数。根据指数函数的特点，当  $a$  和  $\lambda$  值增大时，曲线下下降趋势加快，对应的下降斜率就明显增大。为了避免节点暂时性出现错误的情况，在给出判定阈值  $min\_error$  后，参数取值既要考虑可信度下降的速度，又要考虑迭代步骤的长度。实验表明，当设定  $min\_error=0.18$  时，选取  $\lambda = 0.558$  和  $a = 3.426$  比较合适。

### 4 仿真结果及分析

本节通过模拟仿真实验对 EFSA 的性能进行评价。采用网络仿真软件 NS(network simulator)-2<sup>[17]</sup>作为仿真平台，在进行的仿真实验中，节点被随机分布在一个 1 000m×1 000m 方形区域，选择配置的节点数量在 100~500 个之间变化。MAC 层协议采用 IEEE 802.11 协议。按照 EFSA 机制运行实验，出错节点随机选择并注入出错概率，事件随机选择产生位置。簇头节点收集的数据通过路由协议 AODV 传送到 sink 节点。以下的每个实验都运行 100 次，取其中的平均值作为最终的结果。表 1 给出了实验的参数设置。

参数	数值
每次仿真时间/m	900
数据聚集集合半径 $E\_tole$	0.6~1.0
无线传输半径/m	20~28
无线感知半径/m	18~22
节点初始能量/J	3
传输能量/mJ	14.88
接收能量/mJ	10.25
节点错误率 $p$	0.02~0.2

#### 4.1 数据聚集容错机制 EFSA 的性能验证

数据聚集操作的首要目标是正确地获得事件值，也是衡量数据聚集容错机制 EFSA 性能的重要因素。首先，研究 EFSA 在不同环境情况下，获取事件值与真实值的相对误差变化情况。

如图 3 所示，随着运行时间的增长，EFSA 获得的相对误差会快速下降，原因是错误节点在多次迭代后可信度快速变小，直至识别为错误节点，减少了对聚集操作的影响。另外可以看出，对于不同的网络规模，具有 200 个节点的网络产生的相对误差在 185s 左右迅速减少，而具有 400 个节点的网络产生的相对误差在 270s 才加速减少，但是 600s 后

得到的相对误差比前者低，原因在于高密度的网络容错迭代所需的时间较长。

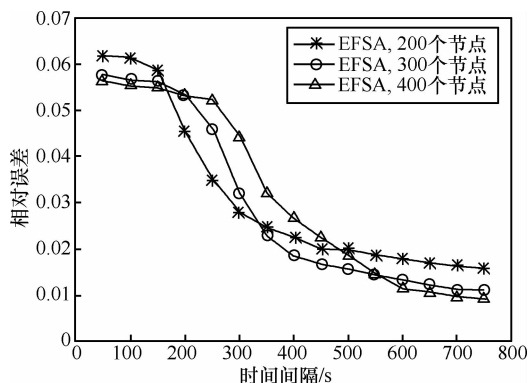


图 3 EFSA 在不同运行时间下的相对误差

接着研究了在不同的网络规模情况下，EFSA 的相对误差比较。如图 4 所示，随着网络规模的扩大，EFSA 机制对应的相对误差都有不同的增长。但是，对于不同的节点出错比例，相对误差的区间变化不大，基本在 0.02~0.03 之间波动，幅度保持相对稳定。图 3 和图 4 的实验结果表明，事件簇使用 EFSA 机制获得的事件值随着实验时间的增加相对误差会逐渐减少并保持相对稳定，特别是在较高密度的无线传感器网络中能获得更好的事件值。

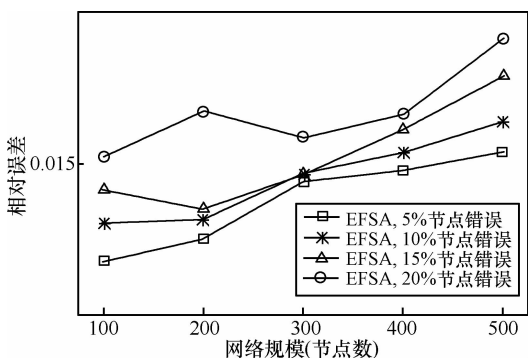


图 4 EFSA 在不同的网络规模下的相对误差

此外，研究了在相同的网络规模下，3 种不同算法得到的事件值误差状况。从图 5 可以看出，错误节点比例增加时，相对误差都呈现增加的趋势。但是，如果固定错误节点比例，Tree 算法比 Gossip 和 EFSA 机制的相对误差要高 2 倍以上，原因在于 Tree 算法没有任何的数据容错功能。Gossip 算法通过随机收敛具有一定的容错能力，因此 Gossip 算法具有的相对误差比 Tree 方法要低。此外，EFSA 机制在出错节点的比例上升时，对应的相对误差只是很缓慢的增长，相对误差比其他的算法都要低且稳

定，这归功于它的容错策略的作用。图 5 的实验结果表明，EFSA 机制对比 Tree 算法和 Gossip 算法能获得更精确的事件值。

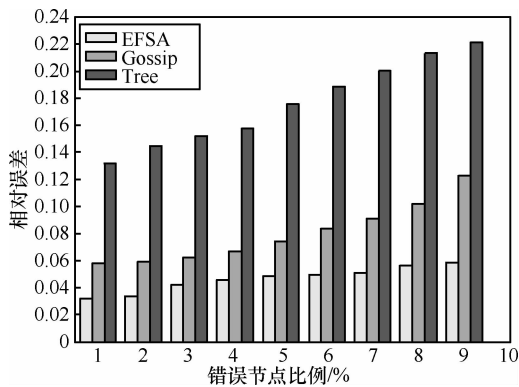


图 5 不同算法在不同错误节点比例下的相对误差

### 4.2 错误节点的检测

仿真实验研究了事件簇错误节点检测的正确率及误报率，它们是衡量容错性能另一个重要指标。假设网络的规模固定，利用节点的可信度来判断节点是否出错。如图 6 所示，在相同的传感器节点出错率的前提下，高密度情况的检测正确性普遍比低密度情况下高 20% 以上，特别是在密度 10 情况下，检测的正确性最低，而密度高于 25 的情况下，检测的正确性比较高，原因在于使用 EFSA 机制时，过低的密度使参加容错过程的数据比较少，很少的错误数据就会影响最终聚集的结果，也不利于错误节点的判断。另外，当密度不变时，随着节点出错概率  $p$  的增大，检测的正确性都会降低，但高密度的情况降低要缓慢些。对于 EFSA 数据聚集容错机制的误报率，也出现相同的情况，从图 7 可知，如果节点密度超过 20 以后，即使节点出错概率的增加，误报率增加的幅度也能保持平稳。

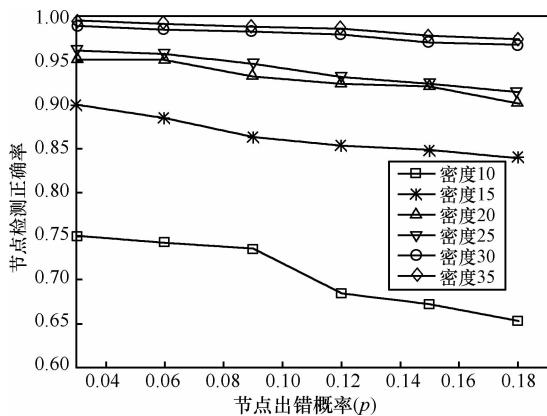


图 6 不同节点错误概率的检测正确率

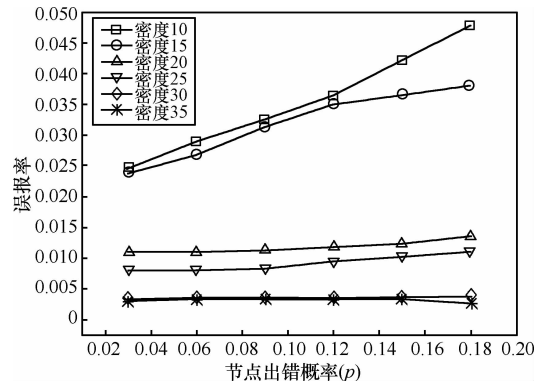


图 7 不同节点错误概率的误报率

图 6 和图 7 的实验表明，数据聚集容错机制 EFSA 在密度较高的情况下具有很好的错误节点的检测正确性能。

### 4.3 事件簇的能量消耗性能

研究数据聚集容错机制 EFSA 的能量消耗性能，因为网络中能量消耗最大部分是数据的传输，只需要考虑节点之间传送的数据分组数量。在固定网络的节点数为 400 时，研究事件簇使用 3 种不同策略产生的数据分组数量对比。如图 8 所示，当节点出错概率  $p$  较低时，Gossip 和 Tree 算法的数据分组传输数量比 EFSA 机制少 25%，原因在于两者都没有事件簇生成过程，数据分组传输的数量比 EFSA 机制要少。但是，当节点出错概率不断提高以后，Tree 算法为了对出错节点采用重构生成树的方式，大大提高了数据分组交换的数量；此外，Gossip 算法也需要重复对出错节点进行数据交换达到数据收敛的目的，造成传输数据分组的数量也有所提高。EFSA 机制只产生相对固定数量的信息分组交换，所以节点出错概率的提高不会额外增加数据分组数量，分组的传输数量基本保持稳定。图 8 的实验表明，事件簇的 EFSA 机制能量消耗波动不大，但是，在节点出错概率较小时能量消耗相对较高。

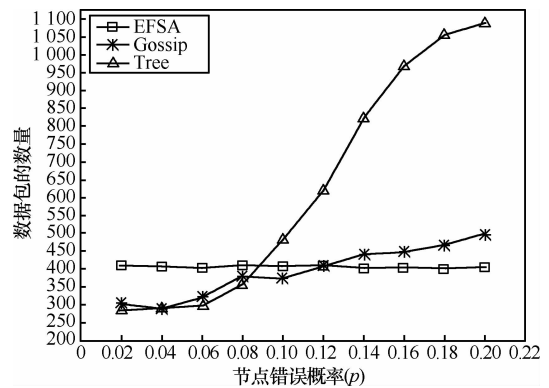


图 8 不同算法的数据分组交换数量

## 5 结束语

在事件簇应用背景下,重点考虑了事件簇在出现错误节点时获得聚集数据、错误影响削减及出错节点确定的问题,提出了事件簇的数据聚集容错机制,研究了基于  $k$ -means 算法的聚集数据获取方法,给出了根据节点可信度的容错及出错节点确定方案。实验结果证明给出的机制能有效地检测出错误节点,并具有可靠性及高效性。然而,在分布式构造事件簇的基础上,寻求数据容错的分布式处理策略是下一步的工作。

### 参考文献:

- [1] PARADIS L, HAN Q. A survey of fault management in wireless sensor networks[J]. ACM Journal of Network and System Management, 2007, 15(2): 171-190.
- [2] 李建中, 李金宝, 石胜飞. 无线传感器网络及其数据管理的概念、问题与进展[J]. 软件学报, 2003, 14(10): 1717-1727.  
LI J Z, LI J B, SHI S F. Concepts, issues and advance of sensor networks and data management of sensor networks[J]. Journal of Software, 2003, 14(10): 1717-1727.
- [3] QUANG V T, MIYOSHI T. Adaptive routing protocol with energy efficiency and event clustering for wireless sensor networks[J]. IEICE Trans Commun, 2008, 91(9): 2785-2805.
- [4] KRISHNAMACHARI B, ESTRIN D, WICKER S. The impact of data aggregation in wireless sensor networks[A]. DCS'02[C]. Washington, 2002. 575-578.
- [5] CHEN H F, MINENO H, MIZUNO T. Adaptive data aggregation scheme in clustered wireless sensor networks[J]. Computer Communications, 2008, 31(15): 3579-3585.
- [6] JAMAL N A, RAZA U M, AHMED E K. Data aggregation and routing in wireless sensor networks: optimal and heuristic algorithms[J]. Computer Networks, 2009, 53(7): 945-960.
- [7] DESOVSKI D, LIU Y, CUKIC B. Linear randomized voting algorithm for fault tolerant sensor fusion and the corresponding reliability model[A]. IEEE International Symposium on Systems Engineering[C]. 2005. 153-162.
- [8] NATHY S, GIBBONS P B, SESHANY S. Synopsis diffusion for robust aggregation in sensor networks[A]. SenSys'04[C]. Baltimore, Maryland, 2004. 250-264.
- [9] GOBRIEL S, KHATTAB S, MOSSE D. Fault tolerant aggregation in sensor networks using corrective actions[A]. SECON'06[C]. Hyatt Regency, Reston, 2006. 595-604.
- [10] WANG Y, WU H Y. DFT-MSN: the delay/fault-tolerant mobile sensor network for pervasive information gathering[J]. Mobile Computing, IEEE Transactions on, 2007, 6(9): 1021-1034.
- [11] KEMPE D, DOBRA A, GEHRKE J. Gossip-based computation of aggregate information[A]. IEEE FOCS'03[C]. Cambridge, MA, USA, 2003. 482-491.
- [12] CHEN J Y, PANDURANGAN G, XU D. Robust aggregates computation in wireless sensor networks: distributed randomized algorithms and analysis[A]. IEEE IPSN'05[C]. Los Angeles, California, USA, 2005. 248-355.
- [13] TERPSTRA W W, BEHNEL S, FIEGE L. A peer-to-peer approach to content-based publish/subscribe[A]. ACM DEBS '03[C]. New York, NY, USA, 2003. 1-8.
- [14] HEINZELMAN W B, CHANDRAKASAN A P, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks[J]. IEEE Trans on Wireless Communication, 2002, 1(4): 660-670.
- [15] KRASNIEWSKI M, VARADHARAJAN P, RABELER B. TIBFIT: trust index based fault tolerance for arbitrary data faults in sensor networks[A]. IEEE DSN'05[C]. Yokohama, Japan, 2005. 672-681.
- [16] LIKAS A, VLASSIS N, VERBEEK J J. The global  $k$ -means clustering algorithm [J]. Pattern Recognition, 2003, 36(2): 451-461.
- [17] The network simulator—ns-2[EB/OL]. <http://www.isi.edu/nsnam/ns/>, 2005.

### 作者简介:



肖伟 (1971-), 男, 湖南长沙人, 国防科技大学博士生, 主要研究方向为移动计算技术、智能计算。



徐明 (1964-), 男, 湖南长沙人, 国防科技大学教授、博士生导师, 主要研究方向为移动计算技术、无线网络及其安全。

吕品 (1983-), 男, 山东滨州人, 博士, 主要研究方向为无线网络与移动计算。

余建平 (1979-), 男, 湖南怀化人, 博士, 主要研究方向为无线传感器网络、群体智能。