一种容错的传感器网络数据融合隐私保护算法*

王涛春[1a,1b], 罗永龙[1a,1b], 左开中[1a,1b]

(1. 安徽师范大学 a. 数学计算机科学学院; b. 网络与信息安全工程技术研究中心, 安徽 芜湖 241003)

摘 要: 由于网络通信具有易错的特点,提出一种具有容错性的隐私保护数据聚集算法。通过椭圆曲线加密方案对节点数据进行加密,保证数据的隐私性,采用加同态加密技术,实现了端到端的聚集加密,节省能耗和带宽。同时,构造轨道图拓扑结构,使得每个节点有多个父节点,当节点与主父节点发生链路失效时,其他父节点能够修复聚集值。仿真实验结果显示,提出的方案在适当增加通信量的情况下,具有较高的数据安全性、很好的容错性和高精度的聚集值。

关键词: 无线传感器网络; 隐私保护; 数据聚集; 容错

中图分类号: TP393; TP301.6 **文献标志码**: A **文章编号**: 1001-3695(2014)05-1499-04

doi:10.3969/j.issn.1001-3695.2014.05.052

Fault-tolerant and privacy-preserving data aggregation algorithm in sensor networks

WANG Tao-chun[1a,1b], LUO Yong-long[1a,1b], ZUO Kai-zhong[1a,1b]

(1. *a. College of Mathematics & Computer Science, b. Engineering Technology Research Center on Network & Information Security, Anhui Normal University, Wuhu Anhui* 241003, *China*)

Abstract: Due to the failure-prone nature of communications, this paper proposed a fault-tolerant and privacy-preserving data aggregation algorithm, which employed an elliptic curve cryptography-based additively homomorphic encryption algorithm to offer end-to-end privacy of sensing data, and then savd energy and bandwidth. Meanwhile, construction track graph topology, each node had more than one parent node, so that when a link error in the primary edge, the backup parent could recover the correct aggregate values. The theoretical analysis and simulation results illustrate this scheme has good privacy of data, low communication overhead and high precision aggregate values.

Key words: wireless sensor networks; privacy-preserve; data aggregation; fault-tolerant

0 引言

作为物联网重要组成部分,无线传感器网络(wireless sensor networks, WSNs)已被广泛应用于野外监控、医疗、军事侦察等民用和军事领域。WSNs 是由大量传感器节点组成,具有资源受限、分布式、自组织、多跳和无线通信等特征[1]。特别是传感节点能量通过电池供电,且不可更换性,使得如何节省能耗,延长 WSNs 使用寿命是 WSNs 研究的关键问题。同时,WSNs 是以多跳和无线通信进行信息传输,使得敌手易于捕获和侦听传输的信息。WSNs 面临严重的隐私数据泄露威胁,且 WSNs 在医疗和军事等应用领域对传输的信息具有较高的安全要求,因此研究和解决 WSNs 中的数据隐私保护问题,对于拓展 WSNs 应用领域具有重要的意义。

WSNs 会产生大量的原始数据,且由于网络的密集部署,节点之间感应范围存在大量的重叠,使得这些原始数据存在较大的冗余性,设计高效的数据处理技术来减少数据冗余以节省能耗和延长传感器使用寿命非常有效。而减少冗余一般通过数据聚集技术实现。很多学者提出很多隐私保护的数据聚集算法[2~6]。通常采用加密的方法实现数据的隐私性,文献[2]提出的 SDAP 方案采用分而治之的思想实现安全的数据聚集。文献[3]提出了 CPDA 和 SMART 两种方案,CPDA 通过在数据中的随机数来隐藏真实数据,簇头节点利用多项式的代数性质求解出加聚集结果。SMART 通过分块—混合技术实现隐私保护的数据加聚集。EEHA[4] 方案与 SMART 方案类似,不同之处是 EEHA 只对叶子节点进行切片—混合处理,中间节点不切片,但对自身私有数据、孩子节点的聚集值和来自叶子节点的切片进行聚集操作,使得 EEHA 方案具有更高的效率和精确度。文献[5]提出了加聚集函数,方案的主要思想是 sink 节点与每个传感节点共享一个随机数(密钥),每个节点将本身传感数据加上随机数,其他操作与一般的数据聚集完全一致,sink 节点得到聚集值减去所有随机数后即得到真正的聚集结果。

由于 WSNs 容易发生节点或链路失效,使得通信经常发生错误,造成聚集结果不准确,很多研究者提出了具有容错功能的数据聚集算法。文献[7]利用时空相关性去预测丢失的数值以进行错误修正。文献[8]通过对后备路由进行转播以修复 hop-

收稿日期: 2013-07-09; **修回日期**: 2013-08-03 **基金项目**: 安徽省高校省级自然科学研究项目(KJ2012Z120);安徽师范大学创新基金项目(2011cxjj04); 安徽省高校省级科学研究重点项目(KJ2011A127)

作者简介: 王涛春(1979-),男,安徽无为人,副教授,博士研究生,主要研究方向为无线传感器网络、安全多方计算(wangtc@mail.ahnu.edu.cn);罗永龙(1972-),男,安徽太湖人,教授,博士,主要研究方向为信息安全、可信计算;左开中(1974-),男,安徽宿州人,教授,博士,主要研究方向为信息安全、三值光计算机.

by-hop 错误。文献[9]提出了一种延迟/容错数据聚集方案,利用参数传输率来决定传输数据分组的时机,并通过容错率来确定数据分组的复制次数。文献[10]提出串联通信共享和分散通信共享方案,通过副本—感知的数据聚集容错机制。

基于此,本文提出一种容错的 WSNs 数据融合隐私保护算法(FTPPDA),采用椭圆曲线同态加密方法对数据进行加密,使得方案能够直接对密文进行聚集操作,实现端到端加密机制,较好地实现了数据的隐私性。同时,通过采用类似于文献[10]提出的多父节点机制使得数据传输具有容错性,孩子节点与主父节点发生链路失效等情况下仍能准确将孩子节点数据传输到 sink 节点,极大地提高了数据聚集的精确度。理论分析和实验结果显示,该方案在安全性、通信量、容错性等方面取得了较好的结果。

1 系统模型

1.1 网络模型及攻击模型

WSNs 由大量资源受限的传感节点构成,本文采用轨道图拓扑结构进行数据聚集操作,如图 1 所示。网络有三种类型的节点,即 sink 节点、中间节点和叶子节点。Sink 节点获得聚集结果;中间节点感知数据并对数据进行聚集操作,即负责接收孩子节点传输过来的数据、聚集数据和将聚集结果传输给父节点。数据聚集包括加、平均值、计数、最大值和最小值聚集等,本文只讨论加聚集,因为其他聚集都可转换为加聚集[5]。本文主要考虑以下几种攻击:a)攻击者通过截获无线信息对传输信息进行窃听;b)攻击者捕获一个或多个节点,并能够获得捕获节点内的数据。本文假设被捕获节点严格按照方案执行各种操作,单会保留所有数据以期推导出其他节点数据。同时,假设 sink 节点是不可被捕获的。

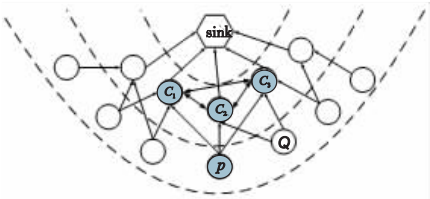

图1 网络拓扑结构

1.2 容错机制

由于传感节点价格低廉,并采用无线通信,使得 WSNs 在进行信息传输过程中,容易发生节点或链路失效等情况,造成聚集结果误差较大,因此为了提高聚集结果的精确性,在进行信息传输过程中给每个节点设置多个父节点,并向多个父节点传输信息以降低由于链路失效等情况造成的误差。具体为将 WSNs 构造成轨道图拓扑结构[10]。如图 1 所示,sink 在轨道 0 上,距 sink 一跳的孩子传感节点在轨道 1 上,以此类推,每个节点被安排在一个轨道上,WSNs 形成具有多条路而不是一个简单生成树的 DAG。

节点有多个父节点,父节点分为主父节点和后备父节点。节点 S_i 将数据广播给父节点,如果主父节点正确接收到节点 S_i 数据后,则主父节点负责对 S_i 进行聚集传输;否则由于节点或链路失效,主父节点没有接收到 S_i 数据,则主父节点向其他父节点广播错误孩子节点信息,节点 S_i 数据由后备父节点按 ID 号由小到大负责聚集传输。例如,节点 P 有主父节点 C_1 和后备父节点 C_2 和 C_3,广播数据 d_p,如果没有发生错误,C_1、C_2 和 C_3 接收到数据,但只有 C_1 负责聚集。假设发生链路错误,节点 C_1 没有收到 d_p,则广播错误信息,C_2 负责聚集 d_p,如果 C_2 也没有正确接收到 d_p,发布错误消息并由 C_3 负责聚集。由此可知,只要 C_1、C_2 和 C_3 中任意一个父节点接收到 P 节点数据,即可正确将孩子节点数据准确地传输到 sink 节点。

1.3 加法同态加密

1)产生密钥 通过给定的安全参数 $\tau \in \mathbb{Z}$ 计算 $g(\tau)$ 得到元组(q_1, q_2, E, n)。E 是椭圆曲线点集合阶为 n 的循环群,其中 $n = q_1 q_2$。从 E 中随机选择两个点(u 和 g),设置 $h = u^{q_2}$,h 为 q_1 阶元。产生公钥为 $P_k = (n, E, g, h)$ 和私钥为 $R_k = q_1$。

2)加密 假设消息空间包含的整数集合为$\{0, 1, \cdots, T\}$,其中 $T < q_2$ 且 T 的位长接近 q_2 的位长。通过公钥 P_k 加密信息 m,选取随机数 $r \leftarrow \{0, 1, \cdots, n-1\}$,并计算出密文 C,即 $C = g^m h^r \in E$。

3)解密 通过私钥 $R_k = q_1$ 对明文 C 进行解密,即 $C^{q_1} = (g^m + h^r)^{q_1} = (g^{q_1})^m$,使得 $\hat{g} = g^{q_1}$,然后基于 $\hat{g}$ 对 C^{q_1} 进行离散 log 运算还原 m 的值[11]。

4)同态加 对两个密文进行聚集,两密文分别为 $C_1 = g^{m_1} h^{r_1}$ 和 $C_2 = g^{m_2} h^{r_2}$,聚集成 C,具体为:$C = C_1 + C_2 = g^{(m_1 + m_2)} + h^{(r_1 + r_2)}$。同态加的具体证明见文献[12]。

2 数据聚集算法

如前所述,数据聚集能够极大地减少通信带宽和能量消耗。高效的数据聚集,特别是需要实现端到端数据隐私在 WSNs 中是非常困难的。同时,传感器网络传输过程中出现数据包丢失情况造成数据聚集结果出现比较大的偏差,需要设计具有容错机制的聚集算法来杜绝或减少这种偏差。本章针对以上情况提出了一种具有容错功能的隐私保护数据聚集算法,其算法的主要思想是:a)采用加法同态加密方案对数据进行聚集,数据隐私达到了端到端的层次;b)通过类多路径方法使得数据聚集具有很高的容错功能。

2.1 FTPPDA 方案

由于 WSNs 通信并不是永远可靠的,会存在节点失效的可能,所以本文提出的数据融合算法不仅具有很好的隐私性,还有很强的容错功能,使得当出现部分链路失效的情况下,数据聚集仍能得到比较精确的结果。方案分四个步骤,具体实现如下:

a)部署密钥。采用 1.3 节所描述的加法同态加密给每个传感节点部署公钥 $P_k = (n, E, g, h)$,sink 节点保留私钥 $R_k = q_1$。

b)轨道图结构的建立。为了实现具有容错性能的数据聚集操作,本文采用类似文献[10]中所描述的方法构造拓扑结构,其中 sink 节点在 0 轨道上,离 sink 一跳的传感节点在 1 轨道上。以此类推,每个节点根据其离 sink 节点跳数确定其所在的轨道。相邻轨道的邻居节点建立父子关系,使得节点可能具有多个父节点或多个孩子节点,如图 1 所示。节点 P 有三个父节点 C_1、C_2 和 C_3,节点 C_2 有两个孩子节点 P 和 Q。同时,每个节点选取的所有父节点是邻居节点[15]。

c)数据聚集及传输。节点分为叶子节点和中间节点。叶子节点通过提前部署的公钥 $P_k = (n, E, g, h)$ 对自身传感数据进行加密,将密文广播给所有父节点。中间节点接收其孩子节

点的信息,采用 1.2 节介绍的容错机制,将自身密文与所有需要负责的孩子节点密文进行同态加聚集操作,并将聚集结果广播给所有父节点。

d)聚集结果。Sink 节点接收到轨道 1 上所有节点的密文聚集结果,通过同态加法运算,并用 $R_k = q_1$ 对其解密得到最终聚集结果,一轮聚集操作结束。

2.2　聚集算法

本文提出的聚集算法的伪代码如算法 1 所述。

算法 1　FTPPDA 算法

构造轨道图拓扑结构;

给所有传感节点 S_i 部署公钥 $S_i \leftarrow P_k = (n, E, g, h)$;

设置等待时间 Δt;

对于每个叶子节点 S_i 的数据执行以下操作

加密:$C_i = \mathrm{Enc}(d_i)_{P_k} = g^{d_i} h^{r_i}$;

将加密后的数据广播给父节点 $P_{ij} \leftarrow C_i$;

对于每个叶子节点 S_i 执行以下操作

在等待时间 Δt 内接收所有孩子节点 Ch_i 数据 $S_i \leftarrow C_{Ch_i}$;

广播孩子节点数据接收情况;

通过 1.2 节介绍的容错机制判断

if S_i 负责聚集 Ch_i 则保留 C_{Ch_i},else 丢弃 C_{Ch_i};

聚集操作:$C_i^{\mathrm{agg}} = C_i + \sum_{Ch_i \in \mathrm{set}_i} C_{\mathrm{Ch}_i}$,其中 set_i 是所有由节点 S_i 负责聚集的孩子节点;

sink 节点执行以下操作

接收轨道 1 上所有节点 S_i 聚集数据 $\mathrm{sink} \leftarrow C_i^{\mathrm{agg}}$;

对所有密文进行同态加操作 $C_{\mathrm{sink}} = \sum_{i \in \mathrm{set}_{\mathrm{sink}}} C_i^{\mathrm{agg}}$,其中 $\mathrm{set}_{\mathrm{sink}}$ 是轨道 1 上节点集合;

解密:$d_{\mathrm{agg}} = \mathrm{Dec}(C_{\mathrm{sink}})_{R_k}$;

结束。

2.3　性能分析

1)安全性

WSNs 最常见的攻击是窃听攻击,对手尝试通过截获到的密文来获取信息,因为本文采用的是椭圆曲线加密方案,依靠大整数因式分解,可以应对该类攻击。另一种攻击是进行明文攻击,即对手通过已知的明文和相应的密文来确定密文信息,本文加密处理依靠随机数,即密文结果是概率性的,所以可以抵抗明文攻击。对手还可以进行主动攻击,即对手捕获传感节点,但由于本文采用非对称加密方案,传感节点只部署了公钥,因此对手不能将聚集的密文进行解密,保证了数据的机密性。因此,对手进行窃听或者捕获传感节点等攻击都不能获取明文信息。

2)通信量

方案通信量主要包括密文数据传输和父节点之间广播孩子节点数据接收情况两个部分。假设密文位长为 q,孩子节点编号用位向量表示,位长为 l_b。对于叶子节点来说只需要传输密文数据即可,假设叶子节点占整个 WSNs 节点数的比例为 p_{leaf}。因此,叶子节点密文传输的通信量为

$$T_{\mathrm{leaf}} = q \times N \times p_{\mathrm{leaf}} \tag{1}$$

中间节点需要接收孩子节点密文信息和其他中间节点广播信息,传输聚集后的密文信息和广播自身孩子节点数据接收情况信息。其通信量为

$$T_{\mathrm{inter}} = N \times (1 - p_{\mathrm{leaf}}) \times ((\mathrm{Num}_{\mathrm{child}} + 1) \times q + \frac{1 - p_{\mathrm{err}}^{\mathrm{Num}_{\mathrm{avgp}}}}{1 - p_{\mathrm{err}}} \times l_b) \tag{2}$$

其中:$\mathrm{Num}_{\mathrm{child}}$ 为中间节点孩子节点数,$\mathrm{Num}_{\mathrm{avgp}}$ 为每个孩子节点具有父节点数,p_{err} 为链路错误率。因此,每个聚集周期主要通信量为 T_{leaf} 与 T_{inter} 之和。

3)精确性

理想情况下,在传输过程中没有数据丢失,则聚集结果完全准确,但由于网络无线通道的冲突性、节点失效性和数据处理的延迟性导致传输信息的丢失,使得聚集结果不完全精确。精确性指的是方案聚集结果与实际的传感数据和之间的比率。为了提高聚集的精确性,本文提出了具有容错性的数据聚集,即将每个节点设置多个父节点,只要其中一个父节点正确接收到节点传感信息,则该节点信息能够准确传输到 sink 节点,极大提高了数据聚集的精确性。其精确性为式(3),其中 $\mathrm{Num}_{\mathrm{par}}^{i}$ 表示节点 s_i 的父节点数。

$$\mathrm{Pre} = \frac{\sum_{i=1}^{N} d_i \times (1 - (p_{\mathrm{err}})^{\mathrm{Num}_{\mathrm{par}}^{i}})}{\sum_{i=1}^{N} d_i} \tag{3}$$

3　实验与分析

为了进一步分析和比较算法的性能,本文在文献[13]的传感器网络模拟器基础上,仿真执行本文提出的 FTPPDA 算法,并对该算法与无容错性和数据机密性基于生成树的数据聚集算法(Tree)、无容错性的隐私保护数据聚集算法(PPDA)、具有容错性但没有隐私保护数据聚集算法(FTDA)在节点密度、链路出错率等参数不同情况下比较其通信量和精确度。实验环境为 Core i3-3220CPU,4 GB 内存;软件环境为 Windows 7 操作系统,VS. NET 2005 和 MATLAB。假设网络覆盖区域 400 × 400 m^2,节点通信半径为 50 m,使用 RC4 对数据进行加密,实验数据集为 Intel Lab Data[14]。

1)N(节点数)对通信量的影响　错误率 $p_{\mathrm{err}} = 0.2$ 时,由图 2 可知,具有容错性的 FTPPDA 和 FTDA 算法所需通信量要比无容错性的 Tree 和 PPDA 高,主要原因是每个节点需要向多个父节点传输数据,同时,父节点之间需要广播信息接收情况。而且,随着 N 的增大,即节点部署密度增加,使得每个节点平均父节点数增多,通信量也相应增加。

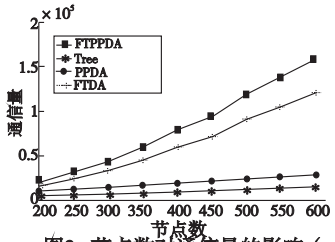

图2　节点数对通信量的影响(p_{err}=0.2)

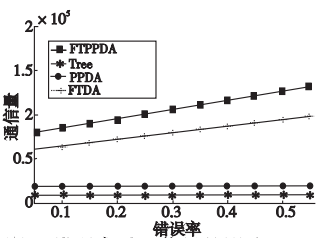

图3　错误率对通信量的影响(N=400)

2)p_{err}(错误率)对通信量的影响　网络节点数 $N = 400$ 时,由图 3 可知,随着链路错误率 p_{err} 变大,FTPPDA 和 FTDA 算法花费的通信量缓慢增加,主要原因是随着 p_{err} 的变大,父节点需要广播信息接收情况的概率增加,但整个增长幅度很小,特

别 p_{err} 在小于 0.3 时。由于 Tree 和 PPDA 不进行错误检测和修正,因此 p_{err} 变化对其通信量没有影响。

3)N(节点数)对聚集精确度的影响　错误率 $p_{err}=0.2$ 时,由图 4 可知,具有容错性的 FTPPDA 和 FTDA 算法的聚集结果的精确度要远高于 Tree 和 PPDA 算法。因为 Tree 和 PPDA 算法,节点在进行信息传输时,当与父节点发生链路等错误,则数据传输失败;而 FTPPDA 和 FTDA 算法,每个节点有多个父节点,且只要能将信息正确传输给任何一个父节点即可,因此聚集结果的精确度有极大的提高。

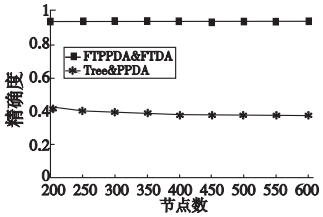

图4　节点数对精确度的影响(p_{err}=0.2)

4)p_{err}(错误率)对聚集精确度的影响　网络节点数 $N=400$ 时,由图 5 可知,随着 p_{err} 的增加,四个算法聚集结果的精确度都随之降低。没有容错性聚集算法的精确度受错误率影响较大,如图 5 所示。当 $p_{err}>0.15$ 时,聚集结果精确度很低,但 FTPPDA 和 FTDA 算法聚集结果精确度降低幅度很小;当 $p_{err}<0.3$ 时,聚集结果精确度和实际结果非常接近。

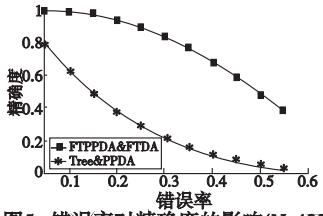

图5　错误率对精确度的影响(N=400)

通过对参数的改变来验证各方案的通信量和聚集精确度。实验结果显示,本文提出的 FTPPDA 算法,不仅保证了传感节点数据隐私性,同时,在适当增加额外通信量的情况下,极大地提高了聚集结果的精确性,使得聚集结果有更高的参考价值。

4　结束语

隐私保护的 WSNs 数据聚集算法已成为新的研究热点。本文提出了具有容错的隐私保护数据聚集算法,采用基于椭圆曲线同态加技术,实现端到端的加密机制,使得传感数据具有较高的隐私性。同时,由于传感节点具有多个父节点,极大减少了由于链路失效造成的数据聚集误差,使得聚集结果具有较高的精确度。本文对方案的通信量、精确度和隐私性进行了理论分析。仿真实验结果显示,该方案在较低通信量的基础上具有较好的隐私性和较高的精确度。

参考文献:

[1] AKYILDIZ I F, SU W, SABKARASUBRAMANIAM Y, *et al*. A survey on sensor networks[J]. IEEE Communications Magazine, 2002, 40(8): 102-114.

[2] YANG Yi, WANG Xin-ran, ZHU Sen-cun, *et al*. SDAP: a secure hop-by-hop data aggregation protocol for sensor networks[J]. ACM Trans on Information and System Security, 2008, 11(4): 32- 43.

[3] HE Wen-bo, LIU Xue, NGUYEN H V, *et al*. PDA: privacy-preserving data aggregation for information collection[J]. ACM Trans on Sensor Networks, 2011, 8(1): 6-27.

[4] LI Hong-juan, LIN Kai, LI Ke-qiu. Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks[J]. Computer Communications, 2011, 34(4): 591-597.

[5] CASTELLUCCIA C, MYKLETUN E, TSUDIK G. Efficient aggregation of encrypted data in wireless sensor networks[C]// Proc of the 2nd Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services. 2005: 109-117.

[6] GROAT M M, HE W B, FORREST S. KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks[C]// Proc of the 30th IEEE International Conference on Computer Communications. 2011: 2024-2032.

[7] MUKHOPADHYAY S, PANIGRAHI D, DEY S. Data aware, low cost error correction for wireless sensor networks[C]//Proc of Wireless Communications and Networking Conference. 2004: 2492-2497.

[8] STANN F, HEIDEMANN J. RMST: reliable data transport in sensor networks[C]// Proc of the 1st International Workshop on Sensor Net Protocols and Applications. 2003: 102-112.

[9] WANG Yu, WU H. DFT-MSN: the delay/fault-tolerant mobile sensor network for pervasive information gathering[J]. IEEE Trans on Mobile Computing, 2007, 6(9): 1021-1034.

[10] GOBRIEL S, KHATTAB S, MOSS D, *et al*. Ridesharing: fault tolerant aggregation in sensor networks using corrective actions[C]//Proc of the 3rd Annual IEEE Communications Society on Sensor, Mesh and Ad hoc Communications and Networks. 2006: 595- 604.

[11] MENEZES A J, Van OORSCHOT P C, VANSTONE S A. Handbook of applied cryptography[M]. [S.l.]: CRC Press, 1997: 128.

[12] BONEH D, GOH E J, NISSIM K. Evaluating 2-DNF formulas on ciphertexts[M]//Theory of Cryptography. Berlin: Springer, 2005: 325-341.

[13] COMAN A, NASCIMENTO M A, SANDER J. A framework for spatio-temporal query processing over wireless sensor networks[C]//Proc of the 30th International Conference on Very Large Data Bases. New York: ACM Press, 2004: 104-110.

[14] SAMUEL M. Intel lab data[EB/OL]. (2004-06). http://db. csail. mit. edu/labdata/labdata. html.

[15] BISWAS S, MORRIS R. Ex: opportunistic multi-hop routing for wireless networks[C]// Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York: ACM Press, 2005: 133-144.

[16] 李洪兵, 熊庆宇, 石为人, 等. 无线传感器网络中网络层故障容错技术研究进展[J]. 计算机应用研究, 2013, 30(7): 1921-1928.