一种低通信量的数据融合隐私保护算法

梁庆庆,杨 庚
(南京邮电大学 计算机学院,江苏 南京 210003)

摘 要:无线传感网中的数据融合技术是降低节点通信量的最为有效的方式之一,而隐私保护是用户数据安全性的要求,有效的数据融合隐私保护算法是无线传感应用的重要研究方向。近年来,出现的一些基于数据分片混合的数据融合隐私保护算法,如 SMART(Slicing-Mix-AggRegaTion),在分片数不小于 3 时可以有效保护数据的安全,但在分片交换阶段网络中数据包过多,数据包容易产生碰撞而丢失。文中提出了一种新的数据融合隐私保护算法 LTPART,它在采用一种安全有效的密钥分配策略的基础上,利用新的数据分片算法,降低了安全通信时数据的通信量。在数据融合阶段,LTPART 为每一层分配固定时间片和浮动时间片,来保证节点数据充分融合及融合的精确性。仿真实验表明,在有效保护数据隐私的前提下,LTPART 要比 SMART($J=3$)少 N(N 为网络中节点的数目)次节点间的通信。

关键词:无线传感网;数据融合;隐私保护
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2013)08-0133-04
doi:10.3969/j.issn.1673-629X.2013.08.034

A Low_traffic Privacy-preserving Aggregation Algorithm

LIANG Qing-qing, YANG Geng
(College of Computer, Nanjing University of Post & Telecommunications, Nanjing 210003, China)

Abstract: Data aggregation mechanism in Wireless Sensor Networks (WSNs) is one of the most efficient methods of reducing data communication overhead. And privacy-preserving is the fundamental security requirement of the users' data. Efficient privacy-preserving aggregation algorithms play an important role in applications of the WSNs. In recent years some privacy-preserving aggregation algorithms based on the slicing in WSNs have been brought up, for instance SMART (Slice-Mix-AggRegaTion), which can efficiently protect the security of data when the number of the slices is not less than 3. But there are too many packets in the network during the slices-mixing, SMART suffers high packet loss ratio arising from the packet collision. In this paper, present a new private data aggregation schema called LTPART in WSNs. Based on a secured and efficient key allocation policy, LTPART reduces data communication overhead with a new data slicing algorithm when the security can be ensured. In the period of data aggregation, LTPART allocates fixed and floating time slices for each layer to ensure the completion and accuracy of data aggregation of sensor nodes. The simulation result shows that LTPART costs N (N is the number of nodes in the network) less than SMART ($J=3$) in terms of communication under the premise of efficient protection of data privacy.

Key words: wireless sensor network; data aggregation; privacy preserving

0 引 言

无线传感网被认为是可以改变世界的十大技术之一,在实际的应用中它主要是对网络中的节点数据进行测量、收集和分析,这其中包括许多敏感数据。例如,在医疗监控系统中,无线传感网被用来收集病人的实时的健康信息。大多数病人不希望自己的健康信息被泄漏,这就需要隐私保护技术来保证这些数据的安全性。

由于同一时间内在同一区域的节点测得的数据具有重复性,传输需要或者是被处理后的数据比传输大量的原始数据更加有意义。另外,数据传输消耗的能量比处理数据多,且无线传感网中无线信道的带宽有

收稿日期:2012-10-23 **修回日期**:2013-01-30 **网络出版时间**:2013-04-22
基金项目:国家自然科学基金资助项目(60873231,60977069);江苏省自然科学基金(BK2009426);江苏省高校自然科学研究重大项目(11KJA520002)
作者简介:梁庆庆(1988-),男,江苏徐州人,硕士研究生,主要研究方向为无线传感器网络中的数据融合与隐私保护;杨 庚,教授,博士生导师,主要研究方向为无线传感器网络、计算机通信与网络、并行与分布式计算、信息安全。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130422.1722.033.html

限,所以对原始数据进行处理来减少通信量是十分有意义的。而数据融合技术就是数据处理的技术之一,它根据需求对数据进行相应的运算,如典型的融合运算 SUM, AVERAGE, COUNT 等,减少数据的总通信量,节省了大量的能量,因此数据融合技术在无线传感网中有十分重要的作用。

数据融合技术的使用可以显著地降低数据通信量,减少能量的消耗;隐私保护可以防止数据的泄漏,是很多应用必须的要求。加密技术可以保证数据的隐私性得到保护。通常采取 end-to-end 的加密方式,即在源节点进行数据加密,然后在目的节点处解密。但数据融合技术需要中间节点对数据进行处理,通常采取 hop-to-hop 的加密方式,即需要通信的一对节点共享一个对称密钥。所以为了既可以使用数据融合技术减少通信量,又可以达到隐私保护的效果,数据融合隐私保护算法的研究被广泛开展。

文中提出了一种针对 SUM 的低通信量的数据融合隐私保护算法 LTPART,它在 SMART[1] 基础上进行了改进,采用了一种基于 Master Device 和密钥分片传输的密钥分配策略[2],在可以保障安全性的同时利用代数运算将节点测量数据分为 2 片,最大程度地减少因分片过多带来的额外通信量。LTPART 采用树形融合树,在数据融合时分配每一层固定时间片和浮动时间片,固定时间片保证分层发送,浮动时间片减少同一层节点发送数据时数据包的碰撞,提高融合的精度。

1 相关工作

现有的研究[1,3~11]已提出一些数据融合隐私保护算法,如 WenBo He 等人提出的基于融合树的隐私保护算法 SMART(Slice-Mix-AggRegaTion)[1],该算法采用分片的方式,将每一个节点的数据固定地分成 J 片,然后将其中的 $J-1$ 片发送给周围的邻居节点,并接收邻居发来的分片,然后将分片组合成一个新的节点数据,这样在不影响数据的融合结果的同时,又保障了每一个节点测量数据的安全性。

在采用密钥池的密钥分配策略时,SMART 只有在 $J \geqslant 3$,节点的隐私数据被暴露的概率才小于 5%[1]。

ESPART[3] 是对 SMART 的一种改进算法,它根据融合树的树形结构的特性,不再固定节点分片的数目,在保证数据隐私性的同时,减少了发送的分片数目,从而减少了数据的通信量和融合时间。

2 系统模型

2.1 网络模型

在文中,无线传感网由连通图 $G(V,E)$ 表示,其中 V 代表网络中的节点,E 表示网络中的通信链路。N 表示网络中的节点数目。

无线传感网中包含了三种类型的节点:QS(Query Server),中间节点,叶子节点。文中只考察有一个 QS 节点的情况。并且网络中每一个节点都可以进行数据的探测,采集,计算,发送和接收。

2.2 数据融合模型

构建一棵以 QS 为根的数据融合树,QS 负责发送融合请求,并得到最终的数据融合结果;中间节点收集它的子节点发来的数据,并与本身采集的数据进行融合后,向上发送给父节点;叶子节点只需采集数据并发送给其父节点。

文中定义的数据融合函数是 $y(t)=f(d_1(t),d_2(t),\cdots,d_N(t))$,$d_i(t)(i=1,\cdots,N)$ 表示节点 i 在 t 时刻采集到的数据。由于很多数据融合的函数,如 count、average 等都可以通过 sum 融合的结果获得,所以文中只以 sum 为研究对象,即 $y(t)=\sum_{i=1}^{N} d_i(t)$。

2.3 密钥分配策略

假设每一个节点都采用文献[2]中介绍的两种密钥:一是与 Master Device 共享一个安全的 pairwise 密钥 M;二是与融合树中的其他节点共享一个对称密钥。第一种密钥用来防止节点的测量数据被其他节点窃取,第二种密钥用来保护数据传输过程中的安全性。这种密钥分配策略的内存使用量和数据无线传输的通信量的规模是 $O(\log N)$[2],N 为无线传感网中节点的数目。另外,节点间通信使用的对称密钥是由其中一个节点随机产生,然后将密钥进行分片,并将每个分片经过不同的节点传输给另一个节点。由于中间传输的节点不能获取密钥的所有分片的值,所以可以保证节点之间密钥的安全性。

3 LTPART:一种低通信量的数据融合隐私保护算法

假设节点 i 测量的数据值为 d_i,d_i 的下界为 $D_{\text{lowerbound}}$,上届为 $D_{\text{uperbound}}$,节点 i 与 Master Device 共享的密钥值为 M_i,节点产生的随机数表示为 $R(R \in (0,1))$。假设在 Tree_Build 和 Data_Mix 时间片内完成融合树的构建和数据的混合,在(FixT+Δt)时间内完成一层数据向上的融合,其中 FixT 是为每一层节点分配的固定时间片,Δt 是为同一层中不同节点分配的浮动时间片,且 Δt 是[0,RandT]的一个随机值。

算法开始:

(1)准备工作。

在 Tree_Bulid 时间内,利用 TAG[12] 算法构建一棵数据融合树。在向下发送 hello 信号的同时,每一个

节点 $i(i=0,1,2,\cdots,N)$ 随机产生一个整数 $\text{pch}_i(2<\text{pch}_i<[D_{\text{lowerbound}}]/2)$ 并向下发送给其子节点,子节点 j 接收来自上层节点发送的 pch_i 后记录为 ppa_i。对根节点 QS(节点号为 0)的 ppa_0 分配一个固定的值。对每一个节点 $i(i=1,2,\cdots,N)$ 随机在 h 跳内选取一个邻居节点集 $S_i(|S_i|\geqslant 1)$,对于稠密的无线传感网,可以令 $h=1$。

(2)数据混合阶段。

在分配的 Data_Mix 时间片内,首先将节点 i 测量的数据 d_i 按照图 1 所示的数据分片算法进行分片,得到 2 个分片 q_i 和 r_i。

然后从 S_i 任意选取一个节点 j,并将 $r_{ij}=R*r_i$ 利用 i 与 j 共享的对称密钥加密后发送给节点 j,此时节点 i 中的数据 $r_i=r_i-r_{ij}$;节点 j 收到加密数据后,利用与 i 共享的密钥进行解密得到 r_{ji},更新 r_j 的值为 r_j+r_{ji}。

```
Function:Slicing;
Input:di;
Output:qi,ri;
Begin:
qi=[di]/ppai;
ri=[di]mod ppai+(di-[di])+Mi;

Randomly generate rdi∈(0,qi);

qi=qi-rdi;
ri=ri+rdi*ppai;
end;
```

图 1　节点 i 的数据分片算法

(3)数据融合阶段。

利用(1)建立的数据融合树,从最底层开始,每一层在(FixT+Δt)的时间内逐层向上融合。在融合的过程中,叶子节点只需将 $d=p*\text{ppa}_r+r$ 值向上传送给父节点即可。对于融合节点 i,记其子节点的数据为 N_i,收到子节点 j 发送来的数据 d_j 后,与自身数据进行如下的融合:

$$d_i=\sum_{j}^{N_i} d_j+\text{ppa}_i*q_i+r_i,$$

然后将 d_i 向上传送给其父节点,并最终在 QS 获得融合结果 d_0。最后,QS 计算原始数据的融合结果 $\text{SUM}=d_0-\sum_{i=0}^{N} M_i$,算法结束。

4　性能分析

利用 TOSSIM 仿真软件对 LTPART,SMART 等算法进行了仿真,具体的网络配置如下:无线对称信道,标准室内环境,背景噪声 -105.0dBm,高斯白噪声为 4dB,节点的数据传输率为 1Mbps,节点的灵敏度为 -108.0dBm。

4.1　数据通信量

LTPART 算法可以分为融合树的构建,分片混合,数据融合三个阶段。融合树构建阶段,和 TAG 算法相同,首先从 QS 发送一个包含随机整数 pch 的 hello 数据包,接收到此数据包的节点将发送节点设置为自己的父节点,并记录 pch 值,然后再继续发送 hello 数据包;在余数混合阶段,每一个节点将 r 的一个随机分片发送给邻居节点,同时接收邻居节点发来的随机分片;在数据融合阶段,每一层的节点须在规定的时间内将自身数据发送给上一层,为简化算法流程,每一层节点分配的固定时间片和浮动时间片相同(FixT=1.5s,RandT=0.5s)。

图 2 说明了 TAG,SMART,ESPART,PART 算法在整个数据融合过程中发送的数据包的个数与网络节点数目基本成正比关系。分别记 TAG,PART,SMART,ESPART 在 1 次数据融合过程中发送的数据包总个数为 $C_T(N)$,$C_L(N)$,$C_S(J,N)$ 和 $C_E(\text{MinDeg},N)$。由图2可知,节点的数量相同时有 $C_T(N)<C_L(N)<C_E(2,N)<C_S(3,N)$。

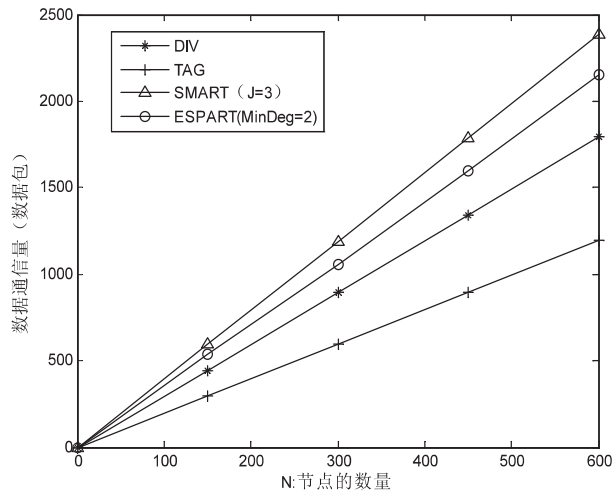

图 2　通信量随节点数目变化图

从理论上对仿真结果进行分析,可知 1 次融合过程中发送的数据包总个数为:

$C_T(N)\approx 2N$,

$C_L(N)\approx 3N$,

$C_S(J,N)\approx (J+1)*N$

由于 ESPART 算法发送数据包的总个数与构建的融合的结构相关,只能从具体的仿真结果中分析得出。

4.2　精确度

由于在每一种数据融合算法中,都不可避免的会因为噪声、碰撞等因素造成最终的融合结果精度的缺失。这里定义精确度为最终的融合结果与节点测量数据的比值。

在分片混合阶段,LTPART 算法中的每个节点只需发送一个时间片,这相比于 SMART($J=3$)在这一

阶段的通信量要小 1 倍,这种低通信量的数据混合可以降低数据包的丢失率。仿真实验表明,这一阶段的丢包率小于 5%。此外,分片数据包的丢失相比于融合阶段数据包的丢失对精度造成的影响要小得多。所以,对融合结果的精度主要考虑融合阶段造成的精度缺失。

为简单起见,每一层分配相同的固定时间片。图 3 显示了固定时间片 FixT 为 1000ms,1500ms,2000ms 时融合的精确度随浮动时间片 RandT 变化的关系图。由图中可以看出,当 RandT 大于等于 400ms 的时候,精确度维持在 90% 左右。

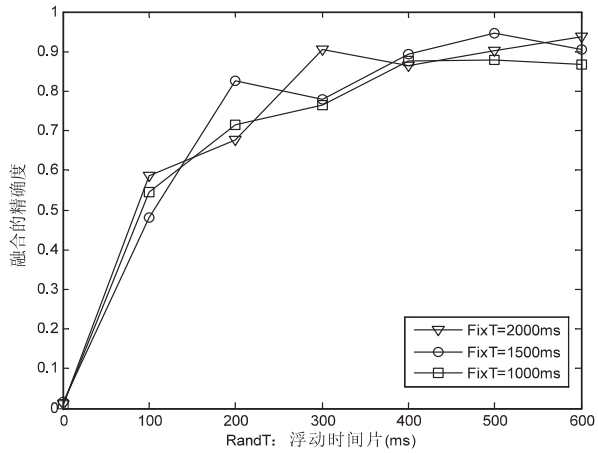

图 3 融合结果的精确性

4.3 计算量

由于不同的节点在整个数据融合过程中的消耗是不同的,将节点分为两类:叶子节点和融合节点(包括中间节点和 QS),数据的计算类型分为代数运算、加密运算和解密运算。记一次代数运算、加密运算、解密运算和依次为 α、β 和 γ[4,5]。假设每个融合节点平均包括 ε 个子节点。

1)SMART 算法的计算量。

首先,对于无线传感网中的所有节点,都需要将测量数据分成 3 片($J=3$)。SMART 采用的是随机分片的方式,假设节点的测量数据是 d,3 个分片依次是 d_1,d_2,d_3。首先要产生两个随机数 rd_1,rd_2,然后依次计算 $d_1 = d * \mathrm{rd}_1$,$d_2 = d * \mathrm{rd}_2$,$d_3 = d - (d_1 + d_2)$,总共需要 4 次代数运算。

接着,每个节点需要将 3 个分片的其中 2 片发送给不同的邻居节点。这时每个节点需要 2 次加密运算。每一个节点平均会接收 2 个数据包,则必须进行 2 次解密运算。此外,还需要 2 次加法运算将 3 个分片相加后得到一个新的节点数据。

最后,叶子节点只需将节点数据发送给父节点即可。这时只需要一次加密运算。而融合节点相比于叶子节点还需要 ε 次解密和 ε 次代数运算。

平均每个叶子节点的计算量为:

$$\mathrm{SC}_{\mathrm{leaf}} = 6\alpha + 3\beta + 2\gamma$$

平均每个融合节点的计算量为:

$$\mathrm{SC}_{\mathrm{agg}} = (6 + \varepsilon)\alpha + 3\beta + (2 + \varepsilon)\gamma$$

所有节点总的计算量为:

$$\mathrm{SC}_{\mathrm{sum}} = (7\alpha + 3\beta + 3\gamma) * N$$

2)LTPART 算法的计算量。

首先每个节点根据图 1 中的算法将测量数据 d 分成 q 和 r,这需要 8 次代数运算。接着将 r 的一个分片发送给邻居节点,这需要 2 次代数运算和 1 次加密运算。每个节点平均接收到一个节点的数据,这需要 1 次解密和 1 次代数运算。然后每个节点通过 $q * \mathrm{ppa} + r$ 获得新的节点数据,这需要 2 次代数运算。最后,叶子节点将数据加密后发送给父节点,这需要 1 次加密运算。而融合节点还需要进行 ε 次解密和 ε 次代数运算。

平均每个叶子节点的计算量为:

$$\mathrm{LC}_{\mathrm{leaf}} = 13\alpha + 2\beta + \gamma$$

平均每个融合节点的计算量为:

$$\mathrm{LC}_{\mathrm{agg}} = (13 + \varepsilon)\alpha + 2\beta + (1 + \varepsilon)\gamma$$

所有节点总的计算量为:

$$\mathrm{LC}_{\mathrm{sum}} = (14\alpha + 2\beta + 2\gamma) * N$$

由上可知,当加密运算和解密运算每次的代数运算大于 4 次时,LTPART 算法的计算量要小于 SMART 的。

4.4 隐私保护性

每一个节点数据通过加入 MD 密钥,使自身数据具有了一定的自我保密性,可以防止数据链路被破解后原始数据泄漏。另外,经过分片混合后,即使是邻居节点也无法获取此节点的原始数据。

5 结束语

文中提出了一种新的数据融合隐私保护方案 LTPART,它采用类除法运算产生随机商和对应余数 2 个分片,并采用一种安全有效的密钥分配策略,在保障节点数据安全的同时,又可以减少数据的通信量,并最终得到精确的融合结果。在仿真的基础上对算法的性能进行了分析。在能保证安全的情况下,LTPART 相比于 SMART($J=3$)要少 N 次节点间的通信。在考虑加密解密运算时,LTPART 实际计算量要小于 SMART。另外,LTPART 可以防御更多种攻击类型。

参考文献:

[1] He W,Liu X,Nguyen H,et al. PDA:privacy-preserving data aggregation in wireless sensor networks[C]//Proc. of the 26th IEEE International Conference on Computer Communications.

(下转第 140 页)

轻度感染——91%

20 ~ 30 岁,女,大于 38.5 度,大于 110 跳/分钟→严重感染——91%

20 ~ 30 岁,男,大于 38.5 度,大于 110 跳/分钟→严重感染——100%

再点击“重点监护者”,这时,“重点监护者”栏中显示的是由上面关联规则判断所得的病情有可能恶化的病人编号和姓名:

0001,李明

0002,晓余

监护人员可以对这两位病人进行重点监护,有效地控制病人的病情。

从以上实验中可知,改进 Apriori 算法在监护中心系统中的实际应用是正确的,并且是有效的、快速的。

4 结束语

传统的多值属性 Apriori 算法先将属性转化为布尔值再使用迭代自连接会产生大量的重复候选集和不必要扫描事务数据库,效率很低。文中提出了一种基于属性值度来找出频繁项集的改进算法。经实验表明,改进 Apriori 算法能够挖掘出持有特征属性形成的强关联规则,这些关联规则表明了生理参数与病情之间的关系,为监护人员更好控制病人病情提供了很好的决策支持。

参考文献:

[1] 胡镜清,刘保延,王永炎. 中医临床个体化诊疗信息特征与数据挖掘技术应用分析[J]. 世界科学技术:中医药现代化,2004,6(1):14-16.

[2] 刘步中. 基于频繁项集挖掘算法的改进与研究[J]. 计算机应用研究,2010,29(2):475-477.

[3] 刘维晓,陈俊丽,屈世富,等. 一种改进的 Apriori 算法[J]. 计算机工程与应用,2011,47(11):149-151.

[4] Gatos B,Mantzaris S,Perantonis S,et al. Automatic page analysis of a digital library from newspaper archives[J]. International Journal of Digital Libraries,2003,3(1):77-84.

[5] Aiello M,Monz C,Todoran L,et al. Document understanding for a broad class of documents[J]. International Journal on Document Analysis and Recognition,2002,5(1):1-16.

[6] 陈立宁,罗 可. Apriori 算法用于频繁子图挖掘的改进方法[J]. 计算机工程与应用,2011,47(10):113-117.

[7] 王德兴,胡学钢,刘晓平,等. 改进购物篮分析的关联规则挖掘算法[J]. 重庆大学学报:自然科学版,2006,29(4):105-107.

[8] Wang Peiji,Shi Lin,Bai Jinniu,et al. Mining Association Rules Based on Apriori Algorithm and Application[C]//Proc. of 2009 International Forum on Computer Science-Technology and Applications. [s. l.]:[s. n.],2009:141-143.

[9] 林郎碟,王灿辉. Apriori 算法在图书推荐服务中的应用与研究[J]. 计算机技术与发展,2011,21(5):22-24.

[10] 张朝晖,陆玉昌,张 钹. 发掘多值属性的关联规则[J]. 软件学报,1998,9(11):801-805.

[11] 高 琰,王台华,郭 帆,等. 应用非迭代 Apriori 算法检测分布式拒绝服务攻击[J]. 计算机应用,2011,31(6):1521-1524.

[12] 汪维清,罗先文,胡继宽. Apriori-Sort 算法研究[J]. 计算机工程与应用,2008,44(36):156-159.

[13] 张宗郁,张亚平,张静远,等. 改进关联规则算法在高校教学管理中的应用[J]. 计算机工程,2012,38(2):75-77.

(上接第 136 页)

Anchorage,AK:[s. n.],2007:2045-2053.

[2] Blaβ E O,Zitterbart M. An Efficient key establishment schema for secure aggregation sensor networks[C]//Proc. of the 2006 ACM Symposium on Information,Computer and Communications Security. New York,USA:[s. n.],2006:303-310.

[3] 杨 庚,王安琪,陈正宇,等. 一种低能耗的数据融合隐私保护算法[J]. 计算机学报,2011,34(5):792-800.

[4] Bista R,Kim H D,Chang J W. A new private data aggregation scheme for wireless sensor networks[C]//Proc. of 10th IEEE International Conference on Computer and Information Technology. Bradford,UK:[s. n.],2010:273-280.

[5] Bista R,Yoo H K,Chang J W. A new sensitive data aggregation scheme for protecting integrity in wireless sensor networks[C]//Proc. of 10th IEEE International Conference on Computer and Information Technology. Bradford,UK:[s. n.],2010:2463-2470.

[6] Groat M M,He W,Forrest S. KIPDA:K-Indistinguishable privacy-preserving data aggregation in wireless sensor networks[C]//Proc. of the 30th IEEE International Conference on Computer Communications. Shanghai,China:[s. n.],2011:2024-2032.

[7] 刘鑫芝. 无线传感器网络安全数据融合算法研究[J]. 计算机与现代化,2010(5):151-155.

[8] 唐 慧,胡向东. 无线传感器网络安全数据融合算法研究[J]. 通信技术,2007,40(12):290-293.

[9] 罗 蔚,胡向东. 无线传感器网络中一种高效的安全数据融合协议[J]. 重庆邮电大学学报(自然科学版),2009,21(1):110-114.

[10] 覃志松,黄延磊. Zigbee 无线传感器网络安全研究及改进[J]. 微计算机信息,2010,26(3-2):54-55.

[11] 邓黎黎,刘才兴. 基于信任的无线传感器网络安全路由研究[J]. 计算机技术与发展,2010,20(6):159-162.

[12] Madden S,Franklin M J,Hellerstein J M. TAG:a tiny aggregation service for ad-hoc sensor networks[C]//Proc. of the 5th Symposium on Operating Systems Design and Implementation. New York,USA:[s. n.],2002:131-146.

一种低通信量的数据融合隐私保护算法

作者： 梁庆庆， 杨庚， LIANG Qing-qing， YANG Geng
作者单位： 南京邮电大学 计算机学院,江苏 南京,210003

刊名： 计算机技术与发展

ISTIC

英文刊名： Computer Technology and Development

年，卷(期)： 2013(8)