基于隐私保护的非线性安全数据融合方案

张 燕,曹晓梅
(南京邮电大学 计算机学院,江苏 南京 210003)

摘 要:数据融合能减少无线传感器节点中传输的冗余信息量,节省传感器的能量消耗和通信带宽。而无线传感器网络由于部署的环境和自身的特点使得在数据融合过程中的隐私信息易受到各种威胁。为了减少信息传输过程中的冗余信息和保护数据融合过程中的隐私信息安全,文中通过对经典安全数据融合方案 ESPDA 进行改进,提出一种基于隐私保护的非线性安全数据融合方案,方案先生成模式码确定上传节点后再进行非线性融合,使其在实现数据融合隐私保护的同时能够在一定程度上减少能耗。仿真结果表明,它不仅能够有效地消除冗余数据的传输,减少能耗,而且还能保证数据安全完整地传输至融合节点。

关键词:无线传感器网络;数据融合;隐私保护;模式码;非线性
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2013)09-0114-05
doi:10.3969/j.issn.1673-629X.2013.09.029

Nonlinear Secure Data Aggregation Scheme Based on Privacy Protection

ZHANG Yan, CAO Xiao-mei
(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Data aggregation can reduce the amount of redundant information transmission in the wireless sensor nodes, save energy consumption and communication bandwidth of the sensor. Due to the deployed environment of wireless sensor networks and its characteristics, the privacy of information during the data aggregation process is vulnerable to a variety of threats. In order to reduce the redundant information in transmission process and protect the privacy information security in information aggregation, propose a nonlinear security data aggregation scheme based on privacy protection through improving the classical secure data aggregation scheme ESPDA, after created pattern-codes to determine the upload node it begins nonlinear aggregation, so that it can protect privacy information during data aggregation process and reduce energy consumption to a certain extent. The simulation results show that it is not only able to effectively eliminate information redundancy, reduce energy consumption, but also to ensure the data is security and integrity transports to the aggregation nodes.

Key words: WSN; data aggregation; privacy protection; pattern-codes; nonlinear

0 引 言

无线传感器网络(Wireless Sensor Netwok, WSN)[1]是能够自主实现数据采集、融合和传输的智能网络,在军事国防、城市管理、工农业控制、抢险救灾、环境监测、危险区域远程控制等诸多领域有着很大的应用前景。由于无线传感器节点密度分布较高,因此各个节点单独传输数据到汇聚节点会使网络存在大量冗余信息,浪费大量能量资源和通信带宽,数据融合[2-3]技术可以被用于解决这个问题。由于 WSN 规模较大,常被部署在无人看守的恶劣环境[4],具有节点能量、计算能力和存储容量等资源受限的特点,使得其在数据采集、传输和融合等环节容易受到各种安全的威胁。

WSN 数据融合隐私保护就是在保证数据融合结果正确的情况下,防止传输的数据被入侵者攻击捕获和解密或是内部的其他可信节点被入侵者俘获来窃取隐私数据信息,或者即使被捕获也能够阻止隐私信息被获取的技术[5]。

收稿日期:2012-12-06 **修回日期**:2013-03-10 **网络出版时间**:2013-05-09
基金项目:国家自然科学基金资助项目(60873231);国家"973"重点基础研究发展计划项目(2011CB302903);江苏高校优势学科建设工程资助项目(yx0020014);青年科学基金项目(61202353)
作者简介:张 燕(1988-),女,硕士研究生,研究方向为计算机通信网与安全;曹晓梅,副教授,研究方向为计算机通信网与安全。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130509.1059.044.html

Cam 提出的能量优先的安全数据融合算法(ESPDA[6])采用生成模式码的方法减少冗余信息[7]的传输,由于簇头节点不进行加解密操作,方案有着一定的安全性,但是协议并没有考虑到数据的隐私性,可能受到攻击而无法保证隐私信息的安全。而用传统的加密方案则有着较大的开销,故文中研究的目的在于对 ESPDA 进行改进,使其在能耗较低的情况下有一定的隐私保护能力。

文中通过对 ESPDA[6]进行改进,提出基于隐私保护的非线性安全数据融合方案(Nonlinear Secure Data Aggregation scheme based on privacy protection, NSDA),使其在数据传输的过程中不仅能够减少冗余信息的传输,减少能耗,而且还能够实现数据融合的隐私保护。

最终实验仿真表明,它不仅能够有效地消除冗余数据的传输,而且还可以保证数据融合过程中的隐私信息在一定程度上得到保护。

1　相关研究

WSN 的安全数据融合要求使用隐私保护技术使得数据隐私信息不被泄露和篡改,同时传感器节点采集到的数据信息能够完整地传输、融合、访问,并且在此过程中,要尽量使方案减少对系统资源的消耗。WSN 中的隐私信息包括位置、时间以及数据方面的隐私信息,现研究的主要是数据隐私的保护。

数据隐私主要是为了防止攻击者通过窃听或者俘获传感器节点的方式来获得或者篡改传输的隐私数据。现对数据隐私方面的保护措施主要有扰乱、查询和加密等保护技术。

W. B. He 和 X. Liu 等人提出了 PDA[8],其中包含两种方案保护数据隐私:SMART 是采用对数据先进行分割再组合的数据分割技术来保护无线传感器节点的隐私数据,在分割的过程中,数据的隐私性得到了很好的保护,而组合的过程能保证数据的完整性;CPDA 采用扰动和分簇的技术,在收集到的数据中添加一定的随机种子和私有随机数,可以扰乱原始数据以达到隐藏真实数据实现隐私保护,传递到簇头节点后,其可以利用多项式的代数性质得到精确的融合值。但是从安全开销分析来看,这两种方案都有着较大的通信计算开销[8]。

杨庚等人设计了 ESPART[9]算法,是对 SMART 方案的一种改进,相比于 SMART,它可以在有效保护数据隐私的前提下,花费少量的时间与数据通信量,得到精确的数据融合结果。

CDA[10]是比较典型的端到端加密机制,传感器节点和基站共享密钥,汇聚节点不需要也无法知道传感器节点的明文信息,可以较好地应对内外部攻击,并且相对于逐跳加密机制中间节点直接对密文进行操作节省了加解密计算开销,减少了时间延迟。但是它们单个叶节点的计算开销增大了,而且只支持 SUM 融合而不支持融合结果的完整性验证。

Cam 等人提出了能量有效的基于模式码的 ESPDA[6]。在 ESPDA 协议中,簇头节点通过自定义模式码的选取阻止传感器节点发送冗余数据,实现数据融合,并使用同态加密体制使得簇头节点执行数据融合时不需要进行加解密操作,一定程度上保证了数据在传输过程中的机密性。

另外,Cam 等还提出了基于参考数据的安全融合协议(SRDA)[11]。通过比较原始采集数据与参考数据,确定差异数据;传感器节点传送差异数据而不是原始数据,从而减少数据传送量。但是由于它们融合后的信息没有进行加密处理等,并不能保证有很好的隐私性,并且数据加解密的过程中也会带来很大的通信开销。

Groat 等人提出的 KIPDA[12]方案是一种典型的非加密保护数据隐私的方案,它在不对节点信息加密的情况下添加伪装数据使数据的隐私性得到一定的保护,实现隐私保护的 MAX/MIN 非线性融合,并可扩展实现隐私保护的 SUM 融合。因为它不需要密钥分配和加解密操作,所在一定程度上节省了节点计算和通信开销,节省了整个网络的能耗。但对于冗余的数据它们不能进行分辨,并且添加伪装数据来扰动原有数据,在一定程度上也加大了计算开销而且此方案的隐私保护能力较弱。

2012 年范永健等人在文献[13]中对近年提出的针对数据聚集、数据查询和访问控制的隐私保护方案所采用的技术做了比较详细的描述,对每种方案的优缺点进行了分析比较。文中提出一种可以消除冗余数据但不使用加密方案的基于隐私保护的 NSDA 方案。

NSDA 方案分两次融合完成:

第一次融合主要是模式码的生成。为了消除冗余的信息,需要将原始数据根据模式生成算法生成模式码[6]。模式码生成后将其与周围节点的模式码进行比较,从而确定要传输的节点数据,至此完成了第一次融合;

第二次融合是根据第一次融合后确定了的进行上传的节点后再将节点中的数据按照一定规律加入一定伪随机数,比较后再上传,最后利用最小或最大原理进行第二次数据融合,由于在此过程中添加的伪随机数及其位置的规律是不按比例的即是非线性的,故把第二次融合又称为非线性融合。

整个方案的网络拓扑图如图 1 所示。

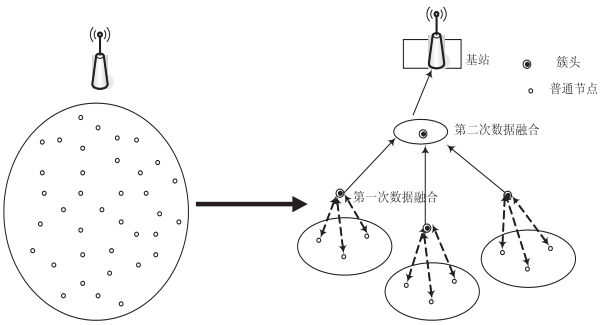

图 1 使用 NSDA 方案的无线传感器网络拓扑图

2 方案实施

2.1 模式码的形成与比较算法

由于 ESPDA 的模式生成算法(PG)是对所有传感器节点适用的,所有此处采用 ESPDA 的模式生成算法。由模式种子生成模式码。传感器节点先从簇头接收到私密的模式种子,每种环境参数的阈值集合决定了数据的间隔值;阈值的数目和间隔值的变化是由用户需求和所部署网络环境中需要的信息精度决定的,然后用模式种子为每个间隔计算判决值,从而生成查询表[6,11]。这里的模式码种子是由任意的数字组成的,并由簇头广播,如表 1 所示。

表 1 判决查找表

阈值	30	50	70	80	90	95	100
区间	0 ~ 30	31 ~ 50	51 ~ 70	71 ~ 80	81 ~ 90	91 ~ 95	96 ~ 100
判定值	5	3	6	1	4	7	2

当节点从环境中感知到数据后,它将会与 PG 算法查询表中定义的间隔值进行比较,从而生成一个判决值。模式码就是将所有采集到的数据参数与判决查找表对照而生成的判决值结合而形成的。

模式生成种子是需要定期进行更新的,以防止攻击者通过长期监听模式码来获取数据信息和操纵数据,从此可以看出模式码技术可以确保融合的安全性和数据的新鲜性。表 2 是采集的数据按照模式生成算法生成的模式码。

表 2 模式码的生成

	Sensor1	Sensor2	Sensor3	Sensor4	Sensor5	Sensor6	Sensor7
D(d1,d2,d3)	(70,25,25)	(58,93,69)	(68,28,30)	(63,24,26)	(56,92,70)	(72,35,29)	(76,32,30)
d1 判决值	6	6	6	6	6	1	1
d2 判决值	5	7	5	5	7	3	3
d3 判决值	5	6	5	5	6	5	5
模式码	655	676	655	655	676	135	135

经过一段时间后,上级节点接收完传感节点的模式码后,整个系统的模式码会存在一定的冗余。将模式码都相同地移动到一个选择集(冗余集)中。例如在表 2 中,节点 1、3 和 4 的模式码相同都是 655,它们就属于同一个冗余集合。每个冗余集中只需要传输一个节点所携带的实际数据信息。选中的节点向其发送数据,另外簇头还会广播一个 ACK 确认信号给其他节点,让这些节点处于休眠状态以节省传输能耗。

故依据上述原则,表 2 中只有节点 1,2,6 这三个节点的这三组数据需要进行上传,去除了一定的冗余信息。

2.2 非线性融合

文中非线性融合的基本思想是:在节点采集到的真实数据中添加一定的伪装数据,构成最终数据信息集,数据集中真实数据的位置是根据基站给出的 GSS 而确定的,再将新构成的数据信息集传递给汇聚节点,由于整个过程中真实数据没有被加密,故汇聚节点能够进行数据融合操作,并且真实数据和伪装数据一起传输,入侵者无法分辨出具体哪些是真实数据和伪装数据,所以可以在不加密的并且隐私信息也得到保护的情况下实现数据融合。另外基站给出的用于确定数据位置的信息 GSS 是不断更新的,使得在不加密情况下数据的隐私性增强了。用 $U^i = \{v_1^i, v_2^i, \cdots, v_n^i\}$ 表示去除冗余信息后的传感节点 i 的消息集,用 $I = \{1,2,\cdots,n\}$ 表示 U^i 位置索引集,有 $|I| = |U^i|$。现介绍新的实现隐私保护的非线性聚集方案。此方案可分为如下 5 个时期:预配置、数据生成、加密、融合、基站处理:

(1)预配置:基站需要选择 I 的子集作为全局秘密信息集(GSS),有 GSS$\subset I$, GSS 为基站私有秘密信息集;GSS 中数据信息表示数据所放位置。如若 $I = |7|$, GSS = {2,3,6} 则表示此信息集总共有 7 位,将数据信息分别放在第二、第三和第六位,剩下的位置用于填充伪装数据。

(2)数据生成:去除冗余信息后剩下的每个节点 i 生成消息集 u^i 并传送至聚集节点。用 $[d_{\min}, d_{\max}]$ 表示感知数据的范围。在 NSS_T^i 对应的 U^i 的位置,放入真实的感知数据 d_i;在 $NSS^i - NSS_T^i$ 对应的 U^i 的位置填充满足以下条件的伪装数据:此时的目的是统计区域内的温度、湿度、气压等的最值,使用 MAX 聚集,伪装数据范围为(0,100)。执行 MAX 融合, N_1、N_2 和 N_6 节点采集的数据分别为上步骤中去除冗余信息后的节点 1,

2,6 的数据信息,按规则随机生成伪装数据填充后, U^1 = {90,70,23,80,70,25,50}、U^2 = {80,58,93,10,20,69,30}、U^6 = {70,72,35,60,50,29,39},如图 2 所示。

(3) 融合:对比每个子节点 j 相应位数的值,形成新的融合节点 i,计算新的 U_l^i 并上传给基站。聚集后新的 $U_l^i = \{U_1^i, U_2^i, \cdots, U_n^i\}$。其中数据项 $U_l^i(l = 1,2,\cdots,n)$ 为原 U^i 与所有子节点上传的 U^j 中相应 l 位置数据的最值。即执行的是 MAX 最值融合,即每个传送到汇聚节点的值都变大了。图 2 示例中计算的新的 U^1 为 {90,72,93,80,70,69,50}。

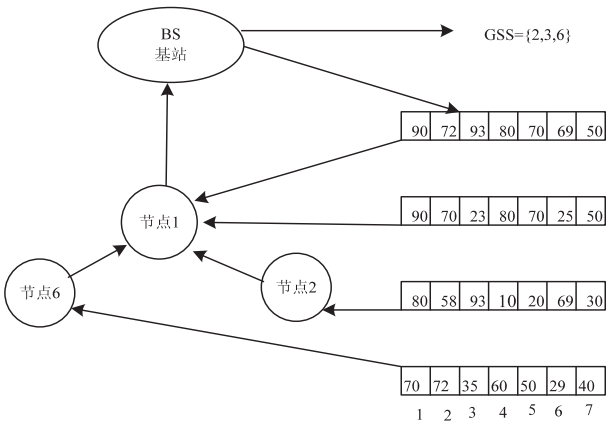

图 2　非线性融合(第二次融合)图

(4)基站处理:基站收到最终汇聚消息 U_l,将其中的 GSS={2,3,6} 所对应位置中的值取出即得到最终融合结果,最终结果为(72,93,69)。这便是所检测环境中温度、湿度、气压的最值,由这个最值在一定程度上可以判断大型温室中一些敏感稀有植物的生存环境是否符合。同时也适用于战场上对敌军的检测通过传感器节点采集的温度等信息确定是否有敌军或者大型设备的靠近或者他们的位置信息。

为保证非线性融合的隐私性,给基站一个触发更新机制,让其对 GSS 的取值不断更新,即节点数据信息存放位置相应地会发生改变,起到了一定隐私保护的作用。

3　仿真与分析

该节对提出的融合方案在能耗方面的性能进行仿真和分析。仿真是在 Ubuntu 下通过编写 TinyOS 应用程序来利用 TOSSIM 的[14]。各个节点动作的控制以及运行流程都通过 TinyOS 应用程序反应出来。在实验中,随机将 1 000 个传感器节点散布在 400 m × 400 m 的区域内,无线信道对称。所有传感器节点具有相同的传输半径 50 m,让其监测环境中的温度、湿度、压强信息均不大于 100,得出所在区域三者的最大值。

图 3 描述了带宽占用率随着感知节点和融合节点间冗余度变化的规律,可以看到,随着感知节点和融合节点冗余度的增大,ESPDA 和 NSDA 的带宽占用率都在不断减小,但是与 ESPDA 协议相比,NSDA 方案的带宽占用率下降更快,即具有更有效的带宽利用率。

图 4 是两种方案能耗的比较,两者首先都要进行模式码的生成和比较,但在随后的操作中 ESPDA 将要传送的信息使用传统的加密方法进行加密后送到基站再解密,而文中的 NSDA 方案是使用非线性的方案,添加一定的伪装数据不加密。图 4 是直接利用根据文献[16]所使用加密方案的能耗和文中的 NSDA 方案所需的能耗进行比较的图。

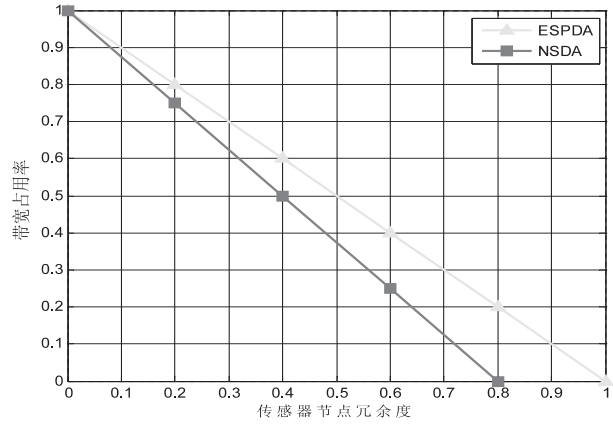

图 3　带宽占用率随感知节点和融合节点冗余度的变化

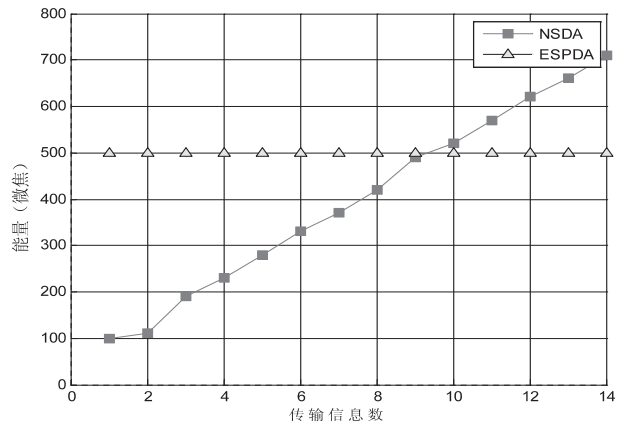

图 4　能耗比较

图 4 中可以看出 NSDA 方案在模式码选择算法后需要传输的节点数为 9 个左右时有着比较低的能耗,这是因为当传输节点较少时,伪装数据相对而言能耗也小点;当传输的节点数超过 9 个时,其伪装数据的加入,这部分的能耗在一定程度上影响了总体方案的能耗。比如文中传输 3 个节点数据时,与加密方案相比能耗就低得多。说明这种方案也存在着一定的局限性,即对传输区域的大小有一定限制。

数据在进行非线性融合前,加入了一定的伪装数据,隐藏了原始数据,并且原始数据存放的位置由于 GSS 的改变而不断改变,使得数据的隐私性得到了很好的保护。另外与传统的通过加密进行隐私保护的方

案相比,文中方案由于在整个过程中不涉及加解密操作,从而节省了加解密过程中带来的时延,即降低了网络时延。

4 结束语

文中提出的 NSDA 方案与 ESPDA 方案相比,有着更有效的带宽利用率,即能够更加有效地消除冗余信息,并且与传统的加密方案相比,在一定程度上有着更低的能耗。而且由于数据在最终的传输过程中加入了一定的伪装数据,隐藏了原有数据,使其隐私性能够得到很好的保护。所以对于数据隐私性有着一定要求,同时又要求能耗低的场所比如说温室稀有植物的培养、战场仪器的监测等这些领域,就可以考虑文中提出的方案。

因为可以通过改变 I 和 GSS 的值来达到获得更强的隐私保护的目的,所以未来的工作主要是找出合适的传输数据位数 I 、GSS 的值;同时为了防止捕获基站信息,可以考虑将基站信息 I 和 GSS 进行加密,防止被攻击者获取。

参考文献:

[1] 沈玉龙,马建峰. 无线传感器网络安全技术概论[M]. 北京:人民邮电出版社,2010.

[2] 回春立,刘 巍,张豫鹤. 无线传感器网络中的数据融合及其能效评估[J]. 计算机应用研究,2008,25(2):546-550.

[3] 吴 凡,胡斌杰. 数据融合算法在 ZigBee 网络中的应用研究[J]. 计算机技术与发展,2011,21(2):226-229.

[4] 曹丛柱,方木云. 无线传感器网络安全问题浅析[J]. 电脑知识与技术,2009,5(16):4161-4163.

[5] Chan H,Perrig A. Security and privacy in sensor networks[J]. IEEE Computer Magazine,2003,36(10):103-105.

[6] Ozdemir S,Cam H. ESPDA:Energy efficient and secure pattern based data aggregation for wireless sensor networks[C]// Proc. of the 2nd IEEE Conference on Sensors. New York: IEEE Society Press,2003.

[7] Curiac D I,Volosencu C,Pescaru D. Redundancy and its applications in wireless sensor networks: A survey[J]. WSEAS Transactions on Computers,2009,8(4):705-714.

[8] He Wenbo,Liu Xue,Nguyen H. PDA:Privacy-preserving data aggregation in wireless sensor networks[C]//Proc. of the 26th IEEE International Conference on Computer Communications. Washington:IEEE Computer Society Press,2007:2045-2053.

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

[10] Girao J,Westhoff D,Schneider M. CDA:Concealed data aggregation for reverse multicast traffic in wireless sensor networks[C]//Proceedings of IEEE International Conference on Communications (ICC). Seoul,Korea:[s. n.],2005:3044-3049.

[11] Sanli H O,Ozdemir S,Cam H. SRDA:Secure reference-based data aggregation protocol for wireless sensor networks[C]// Proceedings of the IEEE VTC Fall Conference. Los Angeles, CA:[s. n.],2004:4650-4654.

[12] Groat M M,He W B,Forrest S. KIPDA:K-Indistinguishable Privacy-preserving Data Aggregation in Wireless Sensor Networks[C]//Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM). Shanghai,China:[s. n.],2011:2024-2032.

[13] 范永健,陈 红,张晓莹. 无线传感器网络数据隐私保护技术[J]. 计算机学报,2012,35(6):1131-1146.

[14] Levis P,Lee N,Welsh M,et al. TOSSIM:Accurate and scalable simulation of entire TinyOS applications[C]//Proceedings of the 1st International Conference on Embedded Networked Sensor Systems. Los Angeles,USA:[s. n.],2003:126-137.

[15] Prasanth G,Ramnath V,Pushkin P,et al. Analyzing and modeling encryption overhead for sensor network nodes[C]//Proc. of the 2nd ACM Intl. Conf. on Wireless Sensor Networks and App.. New York,USA:[s. n.],2003:151-159.

(上接第 109 页)

[J]. 仪器仪表学报,2004,25(4):696-697.

[3] 马腾飞,郑永果,赵卫东. 基于边缘检测与 Hough 变换的车牌字符分割算法[J]. 系统仿真学报,2006,18(Sup):391-392.

[4] Kiryati N,Eldar Y,Bruckstein A M. A Probabilistic Hough Transform[J]. Pattern Recognition,1991,24(4):303-316.

[5] 张兴会,刘 玲,杜升之,等. 车牌照定位及倾斜校正方法研究[J]. 系统工程与电子技术,2004,26(2):237-239.

[6] 孙 楠,刘志文. 一种改进的中文文档图像倾斜检测方法[J]. 计算机仿真,2006,23(9):184-187.

[7] 潘梅森,郭国强. 基于图像矩的车牌号码倾斜校正[J]. 计算机辅助设计与图形学学报,2007,19(8):1041-1045.

[8] 李文举,梁德群,崔连延,等. 一种新的车牌倾斜校正方法[J]. 信息与控制,2004,33(2):231-235.

[9] 路小波,包 明,黄 卫. 基于投影的车牌倾斜检测方法[J]. 交通运输工程与信息学报,2004,4(2):10-15.

[10] 王 植,贺赛先. 一种基于 Canny 理论的自适应边缘检测方法[J]. 中国图象图形学报,2004,9(8):957-962.

[11] Canny J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1986,8(6):679-698.

[12] Demigny D,Kamle T. A discrete expression of Canny's criteria for step edge detector performances evaluation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997,19(6):1199-1211.

基于隐私保护的非线性安全数据融合方案

作者： 张燕， 曹晓梅， ZHANG Yan， CAO Xiao-mei
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(9)