第43卷 第6A期
2016年6月

计 算 机 科 学
Computer Science

Vol. 43 No. 6A
June 2016

一种可检测数据完整性的安全数据聚合协议

刘怀进 陈永红 田 辉 王 田 蔡奕侨
(华侨大学计算机科学与技术学院 厦门 361021)

摘 要 在无线传感器网络中如何对传输的聚合数据同时进行数据隐私保护和完整性保护是当前物联网应用中的重要挑战。Ozdemir等人提出的PRDA(Polynomial Regression Based Secure Data Aggregation)协议基于分簇思想并利用多项式性质对聚合数据进行隐私保护,但无法验证数据的完整性。针对PRDA协议的聚合数据可能被篡改或伪造等问题,提出了一种可检测数据完整性的安全数据聚合协议iPRDA。该协议采用多项式函数和数据扰动技术对数据进行隐私保护,通过利用数据之间的关联特性在基站进行完整性检测。实验表明,该方案在不影响数据机密性的条件下,能有效地进行数据完整性检测。

关键词 多项式回归,数据聚合,隐私保护,数据完整性
中图法分类号 TP309 **文献标识码** A

Integrity-checking Security Data Aggregation Protocol

LIU Huai-jin CHEN Yong-hong TIAN Hui WANG Tian CAI Yi-qiao
(Department of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract In wireless sensor network (WSN), how to aggregate data transmission and protect data privacy and integrity is an important challenge in current Internet of things applications. Ozdemir put forward PRDA (Polynomial Regression-based Secure Data Aggregation) protocol based on clustering and properties by using Polynomial of aggregated data privacy protection, but it is unable to validate data integrity. Because aggregated data of PRDA agreement may have been tampered with or counterfeit, this paper proposed a security data aggregation protocol called iPRDA which can detect the data integrity. It uses polynomial functions and data perturbation technology to protect data privacy, and detects data integrity using the link between the data features in base stations. Experiments show that this scheme under the condition of not affecting the data confidentiality, data integrity can be test effectively.

Keywords Polynomial regression, Data aggregation, Privacy protection, Data integrity

1 引言

无线传感器网络中每个节点的能量和资源有限,为了降低节点的能量消耗,延迟网络的生命周期,数据融合技术是一个很好的选择[1]。数据聚合机制在无线网络中通常专注于防止冗余数据的传播,Wang等人[2]提出了一个基于多项式回归的数据聚合算法,其有效地降低了数据冗余。在实际应用中,传感器所处环境并不安全,无线信道可能被窃听,网络中传输的隐私数据也可能在数据融合过程中被篡改。因此,根据无线传感器网络和数据融合的特性,在数据融合操作过程中保证数据隐私和完整性等安全性能面临许多新的挑战[3]。在无线传感器网络数据聚合隐私保护方案中,He等人[4]提出了两种数据聚合隐私保护方案,即CPDA和SMART协议。其中CPDA(Cluster-based Private Data Aggregation)协议基于分簇思想并利用多项式代数性质聚合数据,降低了通信开销,保护了数据的隐私。而SMART(Slice-Mix-Agg-RegaTe)协议通过基于分片技术和可加性的关联属性,对数据进行切片来隐藏原始数据,提高了数据隐私性。随后Westhoff等人[5]提出了基于对称同态加密的数据聚合隐私保护算法,通过同态加密算法实现数据不用解密就可进行聚合,很好地保护了数据的机密性和用户的隐私性。在文献[6]中,Ozdemir等人提出了一种基于多项回归的数据融合隐私保护算法PRDA(Polynomial Regression Based Secure Data Aggregation),通过对隐秘的多项式系数进行融合操作实现隐私保护。但以上文献均不能保证数据的完整性。在无线传感器网络具有完整性验证的隐私保护方案中,Bista等人[7]提出的数据聚合方案通过私密种子进行隐私保护,利用复数域的虚部检测数据的完整性,但安全性较弱,攻击者可以对实数进行篡改而不被检测出来。随后,周强等人[8]提出的数据聚合方案结合了数据扰动技术和分片技术,利用二元数据间的关联性,对数据进

本文受基于数字水印的网络入侵检测的研究(61370007),基于移动雾节点的传感云关键技术研究(61572206),福建省新世纪优秀人才计划项目(2014FJ-NCET-ZR06)资助。

刘怀进(1992—),男,硕士生,主要研究方向为网络信息安全,E-mail:2285316731@qq.com;陈永红(1974—),男,博士,教授,主要研究方向为物联网及安全、云计算与安全、入侵检测、数字水印、大数据安全等;田 辉(1982—),博士,副教授,主要研究方向为网络与信息安全、云计算安全、多媒体内容安全、数字取证、信息隐藏和隐藏通信等;王 田(1982—),男,博士,副教授,主要研究方向为移动计算、物联网及其安全问题、传感云、社交网络、大数据、安卓智能手机应用等;蔡奕侨(1983—),男,博士,讲师,主要研究方向为智能算法及其应用、数据挖掘等。

行了隐私保护和数据完整性验证，但安全性同样较弱，攻击者可以对二元数据值均进行篡改而使最终的融合数据相等却不被检测出来。

针对以上问题，提出一种可检测数据完整性的安全数据聚合算法 iPRDA。该算法利用二元数据之间的关联性对融合数据进行了完整性保护，采用数据扰动技术对数据进行了隐私保护，并引用回归多项式方法减少了数据冗余，提高了通信效率。本文第 2 节解释了系统模型；第 3 节详细介绍本文提出的 iPRDA 算法；第 4 节对 iPRDA 协议进行安全性分析和性能评估；最后对全文进行总结。

2 系统模型

本文采用与文献[6]类似的无线传感器网络模型，如图 1 所示，整个网络被划分成多个簇，每个簇由一个数据聚合器即簇头节点和多个资源有限的传感器节点组成，每个传感器节点可以与簇内的其他节点进行通信，并将感知的数据上传给簇头节点，簇头节点与邻近的簇头进行互相通信，并通过多跳的方式将融合数据发送给基站。

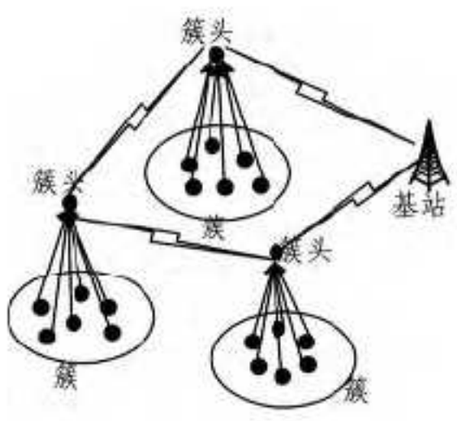

图 1 数据聚合示意图

2.1 数据收集和聚合

在 iPRDA 协议中，数据周期性地收集和聚合。每一次数据聚合过程中，每个数据聚合器通知它的簇成员开始进行数据采集。每个传感器节点接收到消息后采集 n 个感知数据。在一定时间间隔内采集完数据后，数据聚合器要求传感器节点发送它们的感知数据。节点将 n 个感知数据拟合成多项式函数，通过多项式系数代替感知数据发送给数据聚合器。数据聚合器在一定时间内接收完所有的多项式系数后，进行多项式聚合，然后将聚合的多项式发送给基站。

2.2 数据的多项式拟合

为了减少传感器节点传输到数据聚合器的数据总量，假设每个传感器节点预先装入了一个曲线拟合算法，每个传感器节点通过一个 m 次多项式拟合它们的 n 个感知数据。例如

$$f_i(x)=a_{i0}+a_{i1}(x)+a_{i2}x^2+\cdots+a_{im}x^m,n>m$$

请注意，上述公式也可表示成如下：

$$f_i(x)=\sum_{j=0}^{m}a_{ij}x^j,n>m$$

其中，传感器感知的数据集 n 和多项式函数的方次 m 是系统参数，在网络部署前就已经确定好了。通过这些多项式可执行如下加性数据聚合：

$$D_{agg}(x)=\sum_s f_s(x)=\sum_j[(\sum_s a_{sj})]x^j$$

2.3 密钥管理

每个传感器节点与基站共享一个私密种子 $(k_{i,BS},k'_{i,BS})$。每个私密种子在网络部署前已经分配到传感器节点中，基站有所有传感器节点的私密密钥。网络部署之前，在基站和所有的传感器节点安装一个以各自的私密密钥为种子的伪随机数产生器 PRNGs[9]。在后文将解释基站和传感器节点使用 PRNGs 产生随机数。

3 一种可检测数据完整性的安全数据聚合协议

本文基于分簇的思想对数据进行聚合，通过多项式函数和添加随机数进行隐私保护，利用二元数据之间的关联特性进行完整性验证，设计了一种可检测数据完整性的数据聚合隐私保护算法 iPRDA（An Integrity-Protecting Polynomial Regression Based Secure Data Aggregation），它能以相对较低的开销在聚合的同时实现数据隐私保护和完整性验证。

iPRDA 算法由 3 部分组成：节点感知数据和加密，数据传输和聚合，基站接收和验证。

3.1 感知数据和加密

接下来将详细介绍 iPRDA 算法，为了简单其见，只考虑一个簇的情景，如图 2 所示。在这个簇形成和数据聚合器 DA 被选择后，数据聚合器广播一条信息给簇成员，通知数据聚合会话开始。所有的传感器节点获得这个信息后在一定的时间间隔内开始感知数据。每个传感器节点在这次数据聚合会话中执行 n 轮读数并将读数存储到它们的缓冲区内。当所有节点存储完这些读数后，数据聚合器广播另一条信息来请求传感器节点传输数据。接收到数据聚合器的请求消息后，传感器节点首先拟合它的 n 个读数到一个 m 次的多项式（$F_i(x)=\sum_{j=0}^{m}a_{ij}x^j$）中。然后，每个节点 SN_i 使用私密种子（$k_{i,BS}$，$k'_{i,BS}$）通过 PRNGs 产生第 d 次的随机数 $R_i^d=\langle r_i^d,r_i'^d\rangle$，为了保护数据隐私，传感器节点添加 R_i^d 到的 $F_i(x)$ 每个系数中，从而获得隐藏多项式 $eF_i^d(x)=\langle ef_i^d(x),ef_i'^d(x)\rangle$，其中

$$\begin{cases}ef_i^d(x)=\sum_{j=0}^{m}(a_{ij}+R_i^d)x^j\\ef_i'^d(x)=\sum_{j=0}^{m}(a_{ij}+R_i'^d)x^j\end{cases}$$

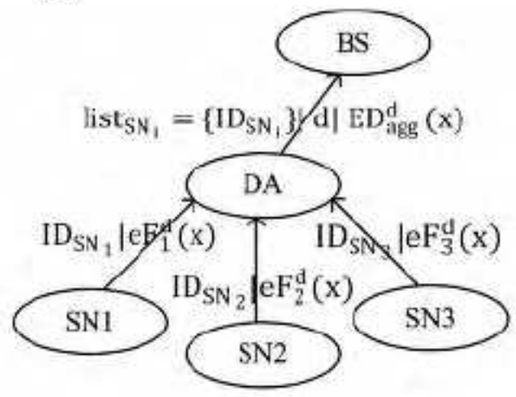

图 2 简单的数据聚合示意图

3.2 数据传输和聚合

节点添加完随机数后发送隐秘系数给数据聚合器，由于随机数只能通过传感器节点和基站共享的密钥产生，因此对隐秘的多项式系数再进行融合操作便实现了隐私保护。每个传感器节点发送隐秘系数和 ID 号给 DA 后，簇头对隐藏多项式进行聚合，获得隐藏聚合多项式 $ED_{agg}^d(x)=\langle EM_{agg}^d(x),EM_{agg}'^d(x)\rangle$，其中

$$\begin{cases}EM_{agg}^d(x)=\sum_s eF_s^d(x)=\sum_j[(\sum_s(a_{sj}+r_s^d))x^j]\\EM_{agg}'^d(x)=\sum_s eF_s'^d(x)=\sum_j[(\sum_s(a_{sj}+r_s'^d))x^j]\end{cases}$$

注意，在数据聚合过程中，DA 不能获得传感器节点的真实数据，因为所有的系数都添加了一个随机数，这样被聚合的数据的隐私性就得到了保证。同时，DA 将第 d 次数据聚合中参与聚合的节点 ID 号添加到一个 IDs 列表中，然后将 IDs 列表随聚合数据发送给基站。其中发送给基站的数据还有 d，符号 d 表示第几次数据聚合。

3.3 基站接收和验证

基站接收隐藏的聚合数据后，根据列表的 ID 号和数据聚合会话数 d，通过私密种子 (k_{iBS}, k'_{iBS}) 产生对应的随机数 R_i^d，然后，基站从隐藏的聚合数据中减去随机数获得聚合数据 $D_{agg}^d(x)=\langle M_{agg}^d(x), M'^d_{agg}(x)\rangle$，其中

$$M_{agg}^d(x)=\sum_s F_s(x)=\sum_j[(\sum_s(a_{sj}+r_s^d-r_s^d))x^j]$$
$$=\sum_j[(\sum_s a_{sj})x^j]$$
$$M'^d_{agg}(x)=\sum_s F_s(x)=\sum_j[(\sum_s(a_{sj}+r_s'^d-r_s'^d))x^j]$$
$$=\sum_j[(\sum_s a_{sj})x^j]$$

基站获得真实的聚合数据后，通过比较 $M_{agg}^d(x)$ 的 $M'^d_{agg}(x)$ 的数值来进行完整性验证。在理想状况下，如果两者相等，则可判断数据完整性未被破坏，接收聚合数据，否则表明完整性被破坏，丢弃聚合数据。

4 性能评估

iPRDA 算法的目的是在保护数据隐私性和进行完整性验证的基础上降低通信开销和获得精确的聚合数据。本节通过仿真实验从隐私保护、通信开销、精确度和完整性 4 个方面分析 iPRDA 的性能，并与 PRDA、iCPDA 等算法进行比较。在 Matlab 仿真环境中执行 iPRDA 算法，实验设定一个在 400m＊400m 的区域内随机分布 200 个传感器节点的无线传感器网络。数据传输的有效范围是 50m，数据传输速率是 1Mbps。采用文献[14]中的通信协议，假设网络被划分成了 N 个簇且每个簇内有 n 个传感器节点，由于感知节点将数据集上传给簇头节点，因此为一跳通信，簇头到基站的平均跳水为 L。假设每个回归多项式有 m 个数据项，每个系数的数据长度为 e，节点的 ID 号数据长度为 l_{id}，则在进行数据融合的过程中，感知节点的通信开销为：

$$T_{i,t}=N\times(l_{id}+e\times m)\times n$$

对于簇头节点，在每个系数长度确定的情况下，聚会的多项式的系数长度由聚合多项式的个数确定，所以簇头节点的通信开销为：

$$T_{DA,t}=N\times(l_{id}+e\times m)\times n\times L$$

因此，在数据融合过程中总的通信开销为：

$$T_{all}=N\times n(l_{id}+e\times m)\times(L+1)$$

4.1 隐私保护分析

本文在 PRDA 协议的基础上进行扩展，在隐私保护方面采用了与 PRDA 类似的机制。因此，对本协议的隐私保护进行分析。一个传感器节点 SN_i 的隐私数据在以下两种情况可能被破解，一是攻击者窃听 SN_i 的传输数据，二是数据聚合器 DA 被妥协。下面分析 iPRDA 的隐私保护性能的这两种情况。

引理 1　一个攻击者不能获取传感器节点的隐私数据，除非它妥协这个传感器节点。

证明：iPRDA 协议通过预先安装在传感器节点和基站的随机数产生器 PRNGs 产生随机数来保护数据隐私[9]。每个传感器节点 SN_i 使用与基站共享的私密种子 (k_{iBS}, k'_{iBS}) 作为随机数产生器 PRNGs 的种子，然后对每次数据聚合会话数 d 产生随机数 $R_i^d=\langle r_i^d, r_i'^d\rangle$。因此，攻击者为了窃听传感器节点 SN_i 的传输数据，必须要有数据聚合会话数 d 和 SN_i 节点与基站共享的私密密钥，密钥在网络部署前就已经分布到各个传感器节点中了，攻击者如果没有妥协这个 SN_i，则不能获得私密密钥。因此，SN_i 的数据隐私是安全的。

引理 2　一个被妥协的数据聚合器 DA 不能获得传感器节点的隐私数据

证明：每个传感器节点发送它的多项式系数给 DA 来代替它的真实数据，为了防止妥协的数据聚合器和窃听者获取隐身数据，每个节点 SN_i 增加一个不同的随机数到它的多项式系数中。由于随机数的产生是基于节点 SN_i 和基站共享的私密种子，为了获得 SN_i 的传输数据，被妥协的 DA 必须要有 SN_i 和 BS 之间的私密种子。由于在网络部署之前私密种子已经分布到每个传感器节点 SN_i 中且 DA 不能获得这个私密种子，因此保证了这个数据聚合过程的隐私泄露问题。

4.2 通信开销

本文是在 PRDA 协议的基础上进行的扩展，在 iPRDA 算法中增加完整性验证功能会带来额外的通信开销，关键问题是这种开销不能太大。使用 Matlab 分别对本文机制和 PRDA 进行仿真实验。首先测试 iPRDA 和 PRDA 在不同的 m/n 值的情况下对通信开销的影响。图 3 表明随着 m/n 值的增加，iPRDA 和 PRDA 的通信开销也在增加。这是因为当 n 值不变时，随着 m 的逐渐增加，从传感器节点传输到数据聚合器的数据通信总量也在增加。同时，因为在 iPRDA 协议中使用了数据之间的关联特性来保证数据的完整性，因此不可避免地要多引入一些通信开销，但实验结果表明其与 PRDA 通信开销差别不大。接下来测试不同簇的大小对通信开销的影响。通过图 4 可以看出，随着簇的逐渐增大，通信开销逐渐减小，这是因为随着簇的逐渐增大，可能在数据聚合器出现拥塞的情况也越严重[10]。

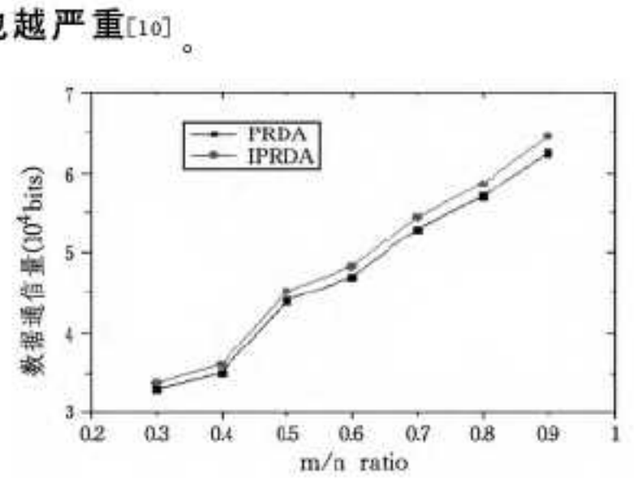

图 3　数据通信量的对比受 m/n 比值的影响

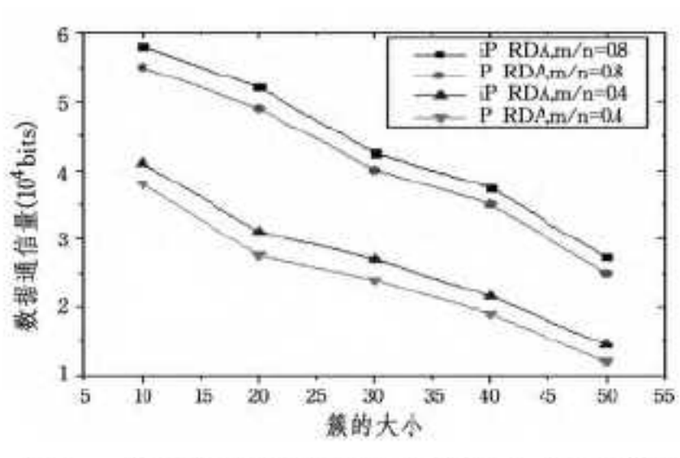

图 4　数据通信量的对比受簇的大小的影响

4.3 数据精确度

iPRDA 的数据聚合的精确度应该与 PRDA 的差别不是很大，同样，也使用不同的 m/n 值和簇的大小来评估本文机制和 PRDA 的数据聚合精确度。图 5 表明，m/n 的比值对数据精确度有一个积极的影响。随着多项式的次数 m 的增加，这个多项式估算原始数据的值就越精确。图 6 表明，簇的大小对数据精确度有一个消极的影响。这是因为，在大型簇中，数据聚合器将出现拥塞，随着拥塞的出现，数据聚合器不能聚合所有传感器节点的传输数据，因此降低了这个数据的精确度。

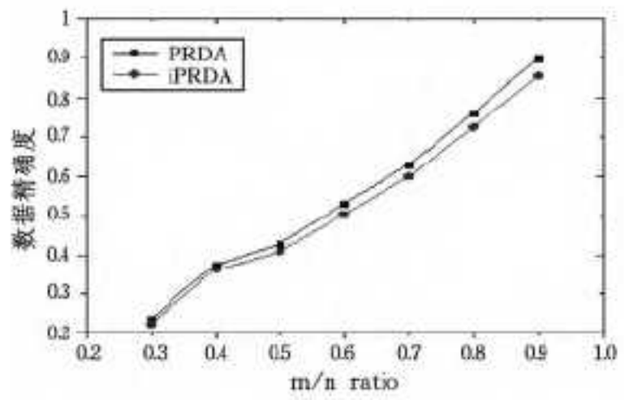

图 5　精确度的比较受 m/n 比值的影响

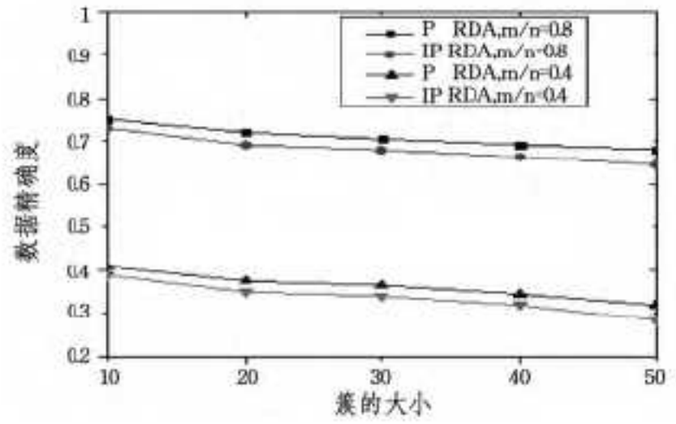

图 6　精确度的比较受簇的大小的影响

4.4　数据完整性

由于 PRDA 没有提供对聚合数据的完整性验证，因此本文的完整性实验与同样具有隐私保护和完整性验证的 iCPDA 进行评估[11]。测评标准为基站最终得到的数据聚合值与实际的数据聚合值之比。1 代表理想状态下数据没有丢失、被篡改的情况，但实际情况中，数据的丢失是难以避免的。如图 7 所示，本文机制在数据完整性保护方面要比 iCPDA 更为完善。传感器节点使用私密种子产生随机数，添加到多项式中不仅隐藏了数据，还使多项式具有关联性，为基站在最后验证数据的完整性提供了很好的保障。

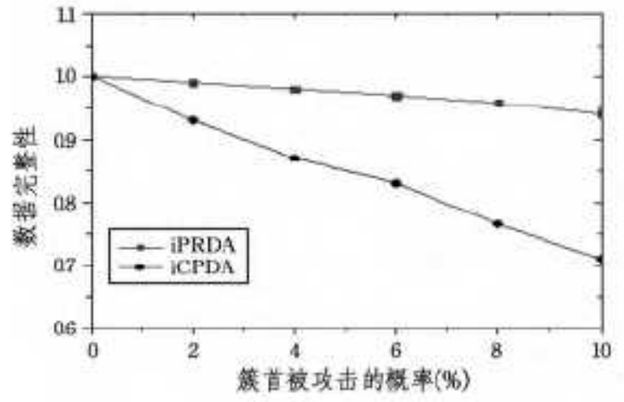

图 7　数据完整性对比

结束语　本文在 PRDA 协议的基础上进行了扩展，提出了一个可以检测数据完整性的安全数据融合算法 iPRDA。iPRDA 采用与 PRDA 协议类似的机制去保护节点的隐私数据。每个传感器节点使用回归多项式去估计它们的真实数据和添加随机数去隐藏相关系数。数据聚合器对隐藏多项式进行数据聚合以保证数据的隐私。通过二元数据之间的关联性对数据进行完整性验证，每个传感器节点使用私密种子产生不同的随机数，然后添加到多项式系数中产生具有关联性的隐私多项式。数据聚合器聚合隐私多项式后发送给基站，基站产生随机数来还原真实的聚合数据。通过仿真实验表明，iPRDA 算法在数据聚合的精确度、数据隐私保护和数据完整性检测等方面均有不错的表现。

iPRDA 算法只是初步工作，还存在不少改进的空间，如进一步降低通信开销和提高精确度等等。

参 考 文 献

[1] Ozdemir S, Xiao Y. Secure data aggregation in wireless sensor networks: A comprehensive overview[J]. Computer Networks, 2009, 53(12): 2022-2037

[2] Wang G, Cao J, Wang H, et al. Polynomial regression for data gathering in environmental monitoring applications[C]// Global Telecommunications Conference, 2007 (GLOBECOM'07). IEEE, 2007: 1307-1311

[3] Madden S, Franklin M J, Hellerstein J M, et al. TAG: A tiny aggregation service for ad-hoc sensor networks[J]. ACM SIGOPS Operating Systems Review, 2002, 36(SI): 131-146

[4] He W, Liu X, Nguyen H, et al. Pda: Privacy-preserving data aggregation in wireless sensor networks[C]// 26th IEEE International Conference on Computer Communications INFOCOM 2007. IEEE, 2007: 2045-2053

[5] Westhoff D, Girao J, Acharya M. Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation[J]. IEEE Transactions on Mobile Computing, 2006, 5(10): 1417-1431

[6] Ozdemir S, Xiao Y. Polynomial regression based secure data aggregation for wireless sensor networks[C]// 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011). IEEE, 2011: 1-5

[7] Bista R, Yoo H K, Chang J W. A new sensitive data aggregation scheme for protecting integrity in wireless sensor networks[C]// 2010 IEEE 10th International Conference on Computer and Information Technology (CIT). IEEE, 2010: 2463-2470

[8] 周强，杨庚，李森，等. 一种可检测数据完整性的隐私数据融合算法[J]. 电子与信息学报，2013，35(6)：1277-1283

[9] Seetharam D, Rhee S. An efficient pseudo random number generator for low-power sensor networks [wireless networks [C]// 29th Annual IEEE International Conference on Local Computer Networks, 2004. IEEE, 2004: 560-562

[10] Ozdemir S. Secure Load Balancing via Hierarchical Data Aggregation in Heterogeneous Sensor Networks[J]. J. Inf. Sci. Eng., 2009, 25(6): 1691-1705

[11] He W, Liu X, Nguyen H, et al. A cluster-based protocol to enforce integrity and preserve privacy in data aggregation[C]// 29th IEEE International Conference on Distributed Computing Systems Workshops, 2009 (ICDCS Workshops'09). IEEE, 2009: 14-19

[12] Du W, Deng J, Han Y S, et al. A pairwise key predistribution scheme for wireless sensor networks[J]. ACM Transactions on Information and System Security (TISSEC), 2005, 8(2): 228-258

[13] Eschenauer L, Gligor V D. A key-management scheme for distributed sensor networks[C]// Proceedings of the 9th ACM Conference on Computer and Communications Security. ACM, 2002: 41-47

[14] Heinzelman W R, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks[C]// Proceedings of the 33rd annual Hawaii international conference on System sciences, 2000. IEEE, 2000: 10

[15] Hu L, Evans D. Secure aggregation for wireless networks[C]// 2003 Symposium on Applications and the Internet Workshops, 2003. IEEE, 2003: 384-391

[16] Mlaih E, Aly S. Secure hop-by-hop aggregation of end-to-end concealed data in wireless sensor networks[C]// INFOCOM Workshops 2008. IEEE, 2008: 1-6