# 一种支持完整性验证的隐私保护直方图融合算法

陈 伟,于 乐,高 迪
(1.南京邮电大学计算机学院,江苏南京 210023;2.江苏省无线传感网高技术研究重点实验室,江苏南京 210003)

**摘 要:** 针对无线传感器网络隐私保护数据融合和完整性验证难以同时兼顾问题,提出一种支持完整性验证的隐私保护直方图融合算法(iPPHA).构建两棵融合树,分别传输融合数据和冗余信息,在基站处对融合结果的完整性进行验证.针对数据包丢失问题,设计了一种 ID 传输方案来提高可靠性.仿真结果显示,算法可以在不明显增加网络资源消耗的前提下,进行完整性验证.改进型 ID 传输方案可节约 70% 的通信开销.

**关键词:** 无线传感器网络; 隐私保护; 数据融合; 完整性验证; ID 传输
**中图分类号:** TP393 **文献标识码:** A **文章编号:** 0372-2112 (2014)11-2268-05
**电子学报 URL:** http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2014.11.021

# A Privacy Preserving Histogram Aggregation Algorithm with Integrity Verification Support

CHEN Wei, YU Le, GAO Di
(1. *College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu* 210023, *China*;
2. *Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing, Jiangsu* 210003, *China*)

**Abstract:** Towards the fact that it is difficult to provide privacy protection and integrity verification simultaneously in data aggregation, we propose a privacy preserving histogram aggregation (iPPHA) algorithm with integrity verification support. It constructs two aggregation trees to transmit aggregate data and redundant information separately. It lets the sink use the redundant information to verify the integrity of the aggregation result. Considering the packet loss problem in WSN, we design an ID transmission scheme flexible bit-map to ensure the reliability of privacy preserving data aggregation and integrity verification. The simulation results show that iPPHA can verify data integrity while protecting data privacy without dramatic resource consumption increase. Flexible bit-map scheme significantly reduces the ID transmission overhead by 70 percent.

**Key words:** wireless sensor network; privacy preservation; data aggregation; integrity verification; ID transmission

## 1 引言

隐私保护数据融合和完整性验证是无线传感器网络(Wireless Sensor Network, WSN)应用中的两大关键技术.考虑到其中数据可能涉及个人隐私,而传感器节点计算、通信能力十分有限[1],对于通信量非常敏感[2],需要进行隐私保护数据融合技术,典型的数据融合算法是 TAG 算法,文献[2]提出了基于分簇和切割数据的两种算法,不能同时降低通信和计算开销.文献[3]中提出的 GP^2S 通过累加一个扰动值,来掩盖直方图表示的真实数据,减小计算开销.考虑到数据在发往基站过程中可能被中间结点篡改,需要进行完整性校验.以较小的资源消耗同时保证数据私密性和完整性成为了近年来研究热点,目前只有少数算法兼顾隐私保护和完整性验证[4],因为数据在上传融合过程中被加密,导致无法进行完整性验证.文献[5]通过簇内节点间的数据交换和广播,防止篡改,但网络规模扩大则通信开销都会快速增长.文献[6]中通过同时建立两棵不同的融合树,分别融合原始数据,在根节点比较两棵融合树的融合结果,保证数据的完整性,但保护能力有限.文献[7]将隐私保护和入侵检测结合起来,但是需要每次设置检测信息.文献[8]在 SMART 算法基础上引入优化因子减小碰撞损失,提高结果精度.

本文提出一种具备完整性验证功能的隐私保护直方图融合算法 iPPHA (Integrity-protecting Privacy Preserving Histogram Aggregation Algorithm).主要采用直方图融合数据,累加扰动值隐藏真实值,另外构建一棵融合树,用于融合汇总少量的冗余信息,在基站处用冗余信息对

收稿日期:2013-07-08;修回日期:2014-06-04;责任编辑:李勇锋
基金项目:国家自然科学基金(No.61202353,No.61272084);江苏省高校自然科学基金(No.12KJB520008);江苏省研究生创新计划(No.CXLX13_464)

数据融合结果进行完整性验证.针对文献[9]提出的 ID 传输问题,设计了一种 Flexible Bit-Map 方法,相比文献[10]所提算法,以更低的带宽消耗传输 ID,反映节点融合参与情况,增强系统鲁棒性.

## 2　系统模型

本文用一张连通图 $G(V,E)$表示整个无线传感器网络.图上点 $V$ 表示传感器节点,线 $E$ 表示通信连接.传感器节点分三种:叶子节点,能感知、上传数据;融合节点,融合子节点数据并上传;基站,融合获得整个网络感知数据.在本方案中融合节点不产生数据,只是汇总和上传数据.基站计算、通信能力远远强于其他节点,并受绝对信任.

系统初始化阶段,将整个无线传感器网络按照位置进行切割、分簇,并在每个簇中选择两个节点作为簇头节点,分别标记成灰色、黑色,剩余节点作为叶子节点,感知到数据后,经过处理发送给所属簇头.颜色相同的簇头节点间相互连接,构成一棵融合树,基站同时作为灰、黑两棵融合树的根.网络中每个节点都分配一个 ID,及一个专享主密钥 $K$,该主密钥只在该节点和基站之间共享.最终得到灰、黑两棵融合树,它们共享全部叶子节点,但是中间的融合节点完全没有交集,如图 1 所示.

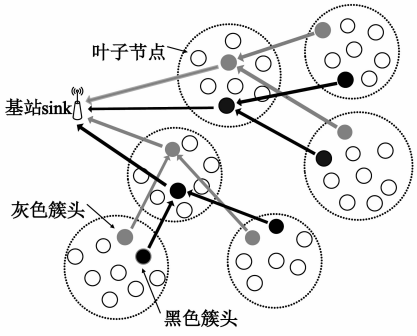


图1　双融合树WSN示意图

## 3　支持完整性验证的隐私保护直方图融合算法:iPPHA

iPPHA 算法分查询请求广播和结果汇总验证两阶段.查询请求广播是指用户发出查询请求后,基站广播发送查询请求包,其中包含本次查询的随机数 $X$,融合树中所有节点收到广播包后层层广播转发,扩散到所有节点.结果汇总验证是本算法核心,包含叶子节点向上层节点发送自身感知数据,汇总到基站,基站解密结果并进行完整性验证这一过程.我们将这一阶段分成隐私保护数据融合、冗余信息融合、完整性验证三个部分依次介绍.

### 3.1　隐私保护数据融合

该部分作用是将所有叶子节点感知到的数据汇总到基站.叶子节点获得某个感知数据,将数据用直方图表示.假定传感器读数范围为$(a,b)$,将这个范围划分成 $K$ 个区间.叶子节点读数为 $t$,所处区间可用下面公式计算:

$$\left\lceil \frac{t-a}{b-a}\cdot K \right\rceil \tag{1}$$

将该区间取值置为 1,其他区间取值置为 0.将这张直方图用二进制数表示,假定整个网络共 $N$ 个叶子节点参与数据融合,因此每个区间的取值在$[0,N]$之间,故可以用$\lceil \log_2 N \rceil$bit 二进制数表示一个区间,整个直方图 $D$ 可以用 $K\cdot\lceil \log_2 N \rceil$bit 表示.在原始数据基础上,为保护数据私密性,叶子节点 $i$ 利用它与基站的主密钥 $K_i$ 产生 $K\cdot\lceil \log_2 N \rceil$bit 扰动值 $P_i$,掩盖真实数据.

$$P_i = \sum_{j=1}^{K} [\, h(K_i \mid X \mid j) \bmod 2^{\lceil \log_2 N \rceil}] \cdot 2^{(j-1)\cdot \log_2 N} \tag{2}$$

其中,$h(\cdot)$表示安全的单向散列函数,$X$ 是本次数据融合的随机数,$\mid$表示将数据连接起来,mod 表示取模.

叶子节点 $i(i=1,2,3,\cdots,N)$将扰动值 $P_i$ 与 $D_i$ 加法取模,将扰动后结果 $R_i$ 上传给该节点对应灰色簇头节点:

$$R_i = (D_i + P_i) \bmod 2^{K\cdot\lceil \log_2 N \rceil} \tag{3}$$

灰色融合节点收到所有子节点传来的加密数据后,将这些值累加并对 $2^{K\cdot\lceil \log_2 N \rceil}$取模,并将结果发往上层灰色节点.这样最终基站将得到整个网络所有叶子节点数据融合结果:

$$R_{\mathrm{SINK}} = \sum_{i=1}^{N} R_i \bmod 2^{K\cdot\lceil \log_2 N \rceil} \tag{4}$$

基站利用所有叶子节点的私钥,生成对应扰动值,$R_{\mathrm{SINK}}$减去各个节点的扰动值 $P_1, P_2, \cdots, P_N$,得到原始数据直方图.

$$\begin{aligned} D_{\mathrm{SINK}} &= \sum_{i=1}^{N} R_i - \sum_{i=1}^{N} P_i \bmod 2^{K\lceil \log_2 N \rceil} \\ &= \sum_{i=1}^{N} D_i \bmod 2^{K\lceil \log_2 N \rceil} \end{aligned} \tag{5}$$

### 3.2　冗余信息融合

该部分目的是将用于完整性校验的冗余信息汇总到基站.叶子节点 $i$ 在将数据 $D_i$ 发送给灰色簇头节点的同时,利用 $D_i$ 生成长度为 $L$ bit 的冗余信息 $C_i$ 并发送给黑色簇头节点.

$$C_i = h(D_i) + h(K_i \mid X) \bmod 2^L \tag{6}$$

黑色簇头节点收到来自子节点的冗余信息后,累加求和,并对 $2^L$ 取模,将结果发往上层节点,最终基站得到所有冗余信息融合结果:

$$C_{SINK} = \sum_{i=1}^{N} h(C_i) \bmod 2^L \tag{7}$$

基站去除融合结果中各个叶子节点的扰动值,得到真实的冗余信息 $C$.

$$C = \sum_{i=1}^{N} C_i - \sum_{i=1}^{N} h(K_i \mid X) \bmod 2^L$$
$$= \sum_{i=1}^{N} h(D_i) \bmod 2^L \tag{8}$$

### 3.3 完整性验证

利用冗余信息可对融合结果完整性进行验证.基站将灰色融合树所得结果进行拆分,还原出各叶子节点原始数据 $D_1, D_2, D_3, \cdots, D_N$,计算用于完整性校验的校验信息 $C'$.基站从黑色融合树获得每个叶子节点的冗余信息融合结果 $C$,将 $C$ 和 $C'$ 进行比较,相同表示数据未被中间融合节点篡改,否则表示受到了污染.

$$C' = \sum_{i=1}^{N} h(D_i) \bmod 2^L \tag{9}$$

## 4 ID 传输方案 Flexible Bit-Map

传感器可能故障,或者发出的数据包丢失,不能参与数据融合,基站需要知道哪些节点参与了此次数据融合,根据这些节点的主密钥才能得到正确的融合结果,因此必须找到一种有效的 ID 传输机制,将整个节点融合参与情况反映给基站[6,9].直接传输 ID 通信开销太大,我们设计了 Flexible Bit-Map,将每个节点的信息压缩到 1bit 以节约带宽.

首先令叶子节点在正常数据后附加自身的 ID,融合节点在正常融合结果后附上自身 ID 和一个 Bit-Map,该 Bit-Map 所占 bit 数等于该融合节点下面叶子节点的数量.每个融合节点在融合子节点传输上来的数据时,对 ID 和 Bit-Map 进行融合处理,得到一个新的 Bit-Map,和正常融合结果一起上传.基站收到最终融合结果后,根据 Bit-Map 查找哪些节点的数据丢失或未参与到融合,处理扰动值时忽略.

为了方便基站将叶子节点与 Bit-Map 的每个 bit 对应,我们规定所有节点按如下方法融合 Bit-Map:假定某融合节点下面有 $m$ 个叶子节点和 $n$ 个融合节点.该节点在构造 Bit-Map 时,用前 $m$ bit 表示它直接连接的叶子节点参与情况,按 ID 从小到大排序.在 $m$ bit 后依次连接 $n$ 个融合节点发送过来的 Bit-Map,分别表示它们下面所属叶子节点的融合参与情况,同样按照融合节点 ID 从小到大排列得到一个新的 Bit-Map.图 2 节点 13 下面共 9 个叶子节点,在构造 Bit-Map 时,用前 3bit 表示自己所属的 3 个叶子节点,后 6bit 表示 3 个子融合节点所属叶子节点参与情况,合计只需 9bit.

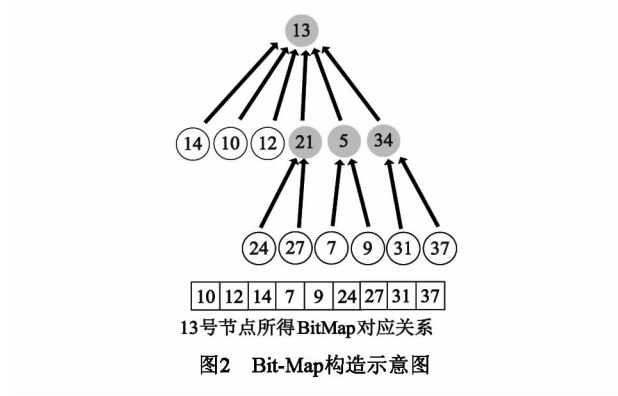


图2 Bit-Map构造示意图

## 5 可行性分析及性能比较

### 5.1 算法可行性

为了验证 iPPHA 算法在无线传感器网络中实际部署的可行性,我们使用 TOSSIM 软件进行仿真,仿真条件如下:

625 个无线传感器节点均匀分布在 50 × 50 的正方形区域中,无线信道对称,标准室内环境,背景噪声为 -105.0dBm,高斯白噪声为 4dBm.选择 50 个节点作为簇头节点,占节点总数的 8%,25 个节点标记为灰色,25 个节点标记为黑色;剩余节点作为叶子节点.仿真 20 次,构建灰、黑两棵融合树.结果显示:50 个簇头中,平均每次有 42.1 个簇头节点成功加入融合树,占总数的 84.3%;575 个叶子节点中,平均有 512.9 个叶子节点可以同时加入两棵融合树,占叶子节点总数的 89.2%;两棵融合树覆盖了整个无线传感器网络,几乎所有叶子节点都同时成功加入到两棵融合树.表 1 统计了各个簇中叶子节点数,68.2% 的簇中叶子节点数在 5 – 35 之间,叶子节点分布较为平均.

表 1 各簇中节点数量分布

| 子节点数 | < 5 | 5 – 15 | 15 – 25 | 25 – 35 | 35 – 45 | > 45 |
|---|---|---|---|---|---|---|
| 簇数 | 12.1% | 31.7% | 19.5% | 17.0% | 14.1% | 5.6% |

### 5.2 算法安全性

被动攻击:iPPHA 在原始直方图基础上累加了扰动值,该扰动值是由单向散列函数产生,攻击者没有该节点的主密钥就无法获得真实数据.各节点利用不同的主密钥产生扰动值,保证即使某些中间节点被攻击者捕获,其他节点的私密性仍然可以得到保护.

主动攻击:基站每次查询给出的随机数可以有效防止重放攻击;由于构建了两棵融合树,这两棵树的中间节点完全没有交集,即使攻击者控制了某个融合节点并篡改融合结果,在基站处通过计算可以发现这种篡改.由于篡改过的数据 hash 值可能跟原数据完全一样,灰色融合树的数据被篡改而不被发现的概率与冗

余信息的长度 $L$ 直接相关,当 $L=5$、8、10 时,基站的漏报率分别为 3.1%、0.39%、0.098%.

### 5.3 算法通信量

iPPHA 算法在 GP^2S 算法的基础上,额外构建融合树传输冗余信息,通信量比 GP^2S 算法略大.冗余信息的长度 $L$ 越大,漏报率越低,安全性越高,额外通信开销越大.图 3 展示了不同网络规模下冗余通信量增长情况.图 4 描述了网络规模增大情况下,冗余通信量在网络通信总量中所占比例变化情况,iPPHA 算法为实现完整性验证所带来的额外开销基本在 10% 以内.

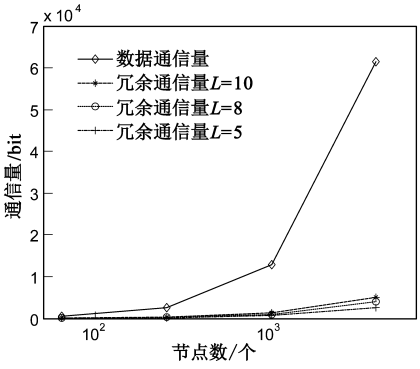


图3　通信量增长

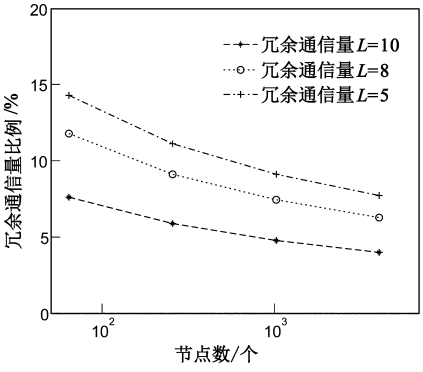


图4　冗余通信量所占比例

图 5 比较了不同网络规模静态分簇网络中 30% 的叶子节点失效时,采用 Flexible Bit-Map 和 Real ID 方案[10]时需要消耗的通信量.相比 Real ID 方案,Flexible Bit-Map 可以节约 70% 的带宽消耗.这主要是由于 Real ID 方案采用定长 Bit-Map 表示 ID,而我们的方案中 Bit-Map 长度可变,因而总体上更优化.

### 5.4 综合性能比较

表 2 将 iPPHA 算法与 iPDA、iCPDA、文献[7]中的算法这三种具备完整性验证功能的算法进行比较.iPDA 对数据切割,本身通信量较大,构建双融合树,则又将通信量增长了 1 倍.iCPDA 层层广播容易造成信息泄露,且时延高.文献[7]中算法主要缺点是需要在每次数据融合前将 $<\tilde{S}_u^i, \tilde{W}_u^i>$ 作为参数发送给融合节点,造成消耗额外通信开销.iPPHA 算法中叶子节点只需发送一个数据包及一个冗余信息包,融合节点只需直接向父节点发送融合结果,因此系统时延和数据包数量明显小于 iPDA 和 iCPDA.采用扰动值隐藏数据,进一步减小加密所需计算量.

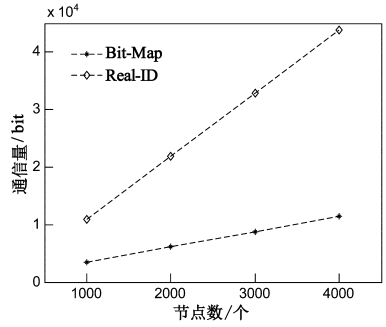


图5　Bit-Map和Real ID方案通信量比较（30%节点失效）

表 2　几种隐私保护数据融合算法性能比较

| 算法性能 | iPDA | iCPDA | 文献[7]中算法 | iPPHA |
|---|---|---|---|---|
| 时延 | 高 | 高 | 一般 | 一般 |
| 计算量 | 低 | 一般 | 一般 | 一般 |
| 通信量 | 高 | 高 | 一般 | 一般 |
| 数据完整性 | 一般 | 高 | 一般 | 高 |
| 隐私保护 | 高 | 一般 | 高 | 高 |

## 6　结论

本文针对无线传感器网络中的完整性保护问题,提出一种改进型算法 iPPHA.在直方图融合基础上,构建另一棵融合树传输冗余信息,以增加系统 5% – 15% 的通信量为代价,在基站处对数据完整性进行验证,防止攻击者对数据进行篡改.同时,针对“有效 ID 传输”问题,设计了 Flexible Bit-Map 方案,在静态分簇网络中,与 Real ID 相比,我们的 ID 传输方案可以节约 70% 的带宽,从而防止部分节点失效影响融合结果.

### 参考文献

[1] Wang Y, Attebury G, Ramamurthy B. A survey of security issues in wireless sensor networks[J]. Communications Surveys & Tutorials, 2006, 8(2): 2 – 23.

[2] He W, Liu X, Nguyen H, et al. Pda: Privacy-preserving data aggregation in wireless sensor networks[A]. Proceedings of the 26th IEEE International Conference on Computer Communications[C]. Anchorage, USA: IEEE, 2007. 2045 – 2053.

[3] Zhang W, Wang C, Feng T. GP^2S: Generic privacy-preservation solutions for approximate aggregation of sensor data[A]. Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications [C]. Hong

Kong, China: IEEE, 2008. 179 – 184.
[4] 许建,杨庚,陈正宇,等. WSN 数据融合中的隐私保护技术研究[J]. 计算机工程,2012,38(15):134 – 138.
Xu Jian, Yang Geng, Chen Zheng-yu, et al. Research of privacy-preserving technology in wireless sensor network data aggregation[J]. Computer Engineering, 2012, 38(15): 134 – 138. (in Chinese)
[5] He W, Liu X, Nguyen H, et al. A cluster-based protocol to enforce integrity and preserve privacy in data aggregation[A]. Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops[C]. Montreal, Canada: IEEE, 2009. 14 – 19.
[6] He W, Nguyen H, Liu X, et al. iPDA: An integrity-protecting private data aggregation scheme for wireless sensor networks [A]. Military Communications Conference [C]. San Diego, USA: IEEE, 2008. 1 – 7.
[7] Wang C, Wang G, Zhang W, et al. Reconciling privacy preservation and intrusion detection in sensory data aggregation[A]. Proceedings of the 30th IEEE International Conference on Computer Communications[C]. Shanghai, China: IEEE, 2011. 336 – 340.
[8] 杨庚,李森,陈正宇,等. 传感器网络中面向隐私保护的高精确度数据融合算法[J]. 计算机学报,2013,36(1):189 – 200.
Yang Geng, Li Sen, Chen Zheng-yu, et al. High accuracy and privacy preserving oriented data aggregation algorithm in sensor networks[J]. Chinese Journal of Computers, 2013, 36(1): 189 – 200. (in Chinese)
[9] Peter S, Piotrowski K, Langendoerfer P. On concealed data aggregation for WSNs[A]. IEEE Consumer Communications and Networking Conference[C]. Las Vegas, Nevada: IEEE, 2007. 192 – 196.
[10] Bista R, Yoo H K, Chang J W. Achieving scalable privacy preserving data aggregation for wireless sensor networks[A]. IEEE 10th International Conference on Computer and Information Technology[C]. Bradford, UK: IEEE, 2010. 1962 – 1969.

## 作者简介

**陈 伟** 男,1979 年 4 月出生,江苏淮安人,博士、副教授,研究方向为网络安全、无线传感器网络等.
E-mail: chenwei@njupt.edu.cn

**于 乐** 男,1990 年 7 月出生,江苏扬州人,硕士研究生,研究方向为无线传感器网络、移动僵尸网络等.
E-mail: yulele08@gmail.com

**高 迪** 男,1990 年 6 月出生,江苏徐州人,硕士研究生,研究方向为无线传感器网络、Android 系统安全.
E-mail: devilkiss6@sina.com