doi:10.11959/j.issn.1000-436x.2015250

传感器网络数据处理中基于隐私向量的隐私保护机制

曾玮妮[1]，林亚平[2]，易叶青[3]，何施茗[2]，陈鹏[1]
（1. 中国船舶重工集团 第 716 研究所，江苏 连云港 222006；2. 湖南大学 信息科学与工程学院，湖南 长沙 410082；
3. 湖南人文科技学院 信息科学与工程系，湖南 娄底 417000）

摘 要：针对传感器网络数据处理中的隐私保护需求，提出了新的分布式机制。构造了隐私向量，并设计了低能耗的隐私向量生成方法及使用方法，从而可有效实现求和、求最值及压缩等各类处理中的数据隐私保护。提出了种子分发算法，保证了隐私向量的安全动态生成。理论分析和仿真实验表明，与已有同类机制相比，新机制不仅能更好地抵御节点俘获攻击，具有更高的隐私保护有效性，且更为能量有效。
关键词：传感器网络；数据处理；隐私保护；同余；隐私向量
中图分类号：TP393; TP309 **文献标识码**：A

Data processing based on the privacy-preserving vector for wireless sensor networks

ZENG Wei-ni[1], LIN Ya-ping[2], YI Ye-qing[3], HE Shi-ming[2], CHEN Peng[1]
(1. The 716th Research Institute, China Shipbuilding Industry Corporation, Lianyungang 222006, China;
2. College of Information Science and Engineering, Hunan University, Changsha 410082, China;
3. Department of Information Science and Engineering, Hunan Institute of Humanities Science and Technology, Loudi 417000, China)

Abstract: To solve the privacy disclosing problem during the data processing phase of wireless sensor networks, a distributed mechanism was proposed. Privacy-preserving vector (vector for short) was constructed. Moreover, lightweight method for vector generation and novel method for using the vector were also presented. Thus, the methods were able to solve the privacy-preserving problem for various data processing funtions such as max/min and data compression efficiently. An algorithm based on data hiding and slice was given and then the vector was able to be generated securely and dynamically. Extensive analyses and experiments show that the mechanism is more robust to node compromise attack and thus can preserve privacy more efficiently. Moreover, the mechanism consumes less power.
Key words: sensor networks; data processing; privacy preserving; congruence; privacy-preserving vector

1 引言

传感器网络（简称传感网）是物联网及信息物理融合系统（CPS，cyber-physical system）的重要组成部分，在医疗卫生、智能家居、国防军事等领域具有广阔的应用前景[1,2]。当传感网应用于医疗卫生及智能家居等领域时，脉搏、心率、水电使用状况等感知数据的暴露可能造成人身安全或道德方面的损失。这要求传感数据不仅要对外部攻击者保持机密性，也需对内部节点保持机密性，即需保证数据的隐私性。只有数据的隐私性得到保证，人们才会普遍接受传感网对其个人信息进行采集，才能真正实现无处不在的感知[3,4]。

由于传感器节点的能量极其有限，传感网通常对采集到的数据进行网内处理（in-network process），再将处理值发送给基站以减少传输能耗。数据处理给传感数据的隐私保护带来了新的挑战：传统加密体系不能在保证数据隐私的同时支持数据

收稿日期：2014-03-26；**修回日期**：2015-07-26
基金项目：国家自然科学基金资助项目（61303045, 61472125, 61472135）
Foundation Item: The National Natural Science Foundation of China (61303045, 61472125, 61472135)

处理；安全多方计算等策略由于开销昂贵同样不适用于传感网[4]。传感网数据处理中的隐私保护研究起步较晚，早期研究主要围绕求和展开[4~12]，而这些工作由于利用了求和的代数特性，不能解决求最值等非线性数据处理中的隐私保护问题。传感网是应用相关的网络，在有些应用场合，需要采用多种数据处理方式。这就需要解决一般数据处理中传感数据的隐私保护问题，目前，针对该问题的研究较少：Zhang 等[13]利用满足特定分布的数扰动传感数据实现隐私保护，然而，该工作不能保证处理结果的精确性，且通信能耗较大；Groat 等[14]提出了基于数据伪装的机制，然而，该机制的隐私保护有效性受限于通信开销且易于遭到数据统计攻击。研究新的轻量级的技术，以解决一般数据处理中传感数据的隐私保护问题至关重要。

为此，本文提出了新的隐私保护机制 PDPV（privacy- preserving data processing scheme based on privacy-preserving vector）。PDPV 不仅适用于求和等线性处理，也适用于求最值及压缩等非线性处理；可以保证处理值的准确性；对节点失效等异常情况具有顽健性；对节点加入具有良好的扩展性；相比已有同类工作，PDPV 在提供更强隐私保护性的同时具有更低能耗。虽然 PDPV 的存储开销高于已有同类机制，然而仍然较小，适合于传感网。

2　相关工作

传感网数据处理中的隐私保护问题受到了学术界的广泛关注，目前已成为研究热点[3]。已有研究成果中，适用于求和等线性处理的成果较多，而适用于求最值等非线性处理的工作还比较有限。

针对求和处理，He 等[4]基于点到点加密技术及数据扰动（perturbation）技术提出了分布式机制 CPDA(cluster-based private data aggregation)；此外，基于点到点加密技术及数据切分技术提出了分布式机制 SMART(slice-mix-aggregate)[4]，其不足在于通信及计算开销均较大。杨庚等[5]在 SMART 的基础上提出了一种低能耗的机制，周强等[6]结合数据扰动和分片技术提出了一种分布式机制，然而，这些机制同样只适合于求和处理。Castelluccia 等[7]、Feng 等[8]基于求和同态加密的思想提出了解决方案：节点与基站共享秘密数，用秘密数隐藏传感数据实现隐私保护，并由基站减去相应秘密数获取和值。这类工作需要所有节点上传数据；否则，需要节点上传其 ID，通信能耗较大。Conti[9]提出的机制中，节点间共享双密钥，通过加减其双密钥隐藏传感数据；通过将隐藏数和传感数据一起求和获取和值。此外，提出了一种隐私保护机制[10]：利用构造的隐私保护元隐藏传感数据实现隐私保护，并通过簇内隐藏后数据的模加运算还原出和值，然而，该机制同样只适合求和处理。

上述机制由于利用了求和的代数特性，仅适用于求和处理。Jung 等[11]基于点到点加密技术及多元多项式提出了一种隐私保护机制，该机制可适用于求和及乘积处理，且节点需要发送多条数据，能耗高。目前，不仅能支持求和，还能支持求最值等非线性处理的隐私保护机制还不多见。Zhang 等[13]利用满足特定分布的数扰动传感数据实现隐私保护，提出了模糊处理机制。其基本思想是：节点将其传感数据映射到直方图区间进行泛化处理，并利用与基站间共享的秘密数实施扰动后再发送给处理节点；处理节点进行求和，并将结果发送给基站；基站减去扰动数，得到所有数据分布情况的直方图，从而可近似获得 max/min、和值及均值等处理值。该机制只能提供近似的处理值，且其通信能耗高。

Groat 等[14]提出了基于数据伪装的解决方案 KIPDA（k-indistinguishable solution to privacy-preserving data aggregation），其基本思想是：节点以明文发送其真实数据及$(|I|-1)$条伪装数据，真实数据在这$|I|$条数据中的位置是事先安排的，因此，处理节点可以进行求最值等处理。然而：1）攻击者通过窃听可获知，真实数据必为$|I|$条数据之一；2）攻击者若俘获了某个节点，则可知真实数据为 k 条数据之一；3）攻击者若俘获了 c 个节点，则可获取所有隐私数据。系统参数 k 及 c 与$|I|$相关且远小于$|I|$。$|I|$越大则隐私保护越有效；而通信开销随$|I|$的增长而迅速增长。此外，若$|I|$固定，则 c 随 k 的增长而迅速降低。因此，KIPDA 的隐私保护有效性受限于通信开销。此外，KIPDA 易于遭到数据统计攻击[14]。为避免攻击，需要基站以单播加密方式更新预置信息，引发较高通信开销和时延。

3　安全模型

3.1　威胁模型及相关假设

同已有文献[4,14]，假设敌方可能实施以下攻击：1）外部窃听攻击；2）节点俘获攻击，且包括

基站(BS)在内的所有节点均可能被俘获，被俘获节点的所有信息均可能被敌方获取；3）串谋攻击，即敌方联合多个被俘获节点获取其他节点的隐私数据。此外，同已有文献[4,14]，假设网络所处的威胁模型为诚实但好奇模型（honest but curious）。该模型中，被俘获节点可能通过多种手段窃取其他节点的隐私数据，但其遵守数据处理等各种预置的协议规范且不篡改数据。

为抵御窃听攻击，在初始化阶段，新提出机制 PDPV 采用对偶密钥（pair-wise key）实现节点对间信息传输的安全性。对偶密钥的管理目前已有较多研究成果，在此基础上假设节点对间对偶密钥具有互异性[15,16]，这意味着获取某对节点间的对偶密钥等同于俘获其中至少一个节点。在数据汇报阶段，PDPV 中簇节点与处理节点间传输的数据实现了隐藏保护，无需加密即可抵御窃听攻击。

3.2 安全目标

数据隐私保护按安全程度可分为 3 个级别[14]，其中第一级别的目标是避免敌方通过窃听获取节点隐私数据，这一级别所能提供的安全性最低。第二级别是使网内节点不能获取其他节点的隐私数据信息，即能一定程度上抵御内部攻击。第三级别的安全目标是使节点数据不会被包括基站在内的任何个体所获取，该级别所能提供的安全性最高，故而也最难实现。在已有的典型工作中，PDA 及 SMART 实现的是诚实但好奇模型下第三级别的隐私保护；KIPDA 所实现的是诚实但好奇模型下第二级别的隐私保护，其基站可以获取节点的隐私数据[14]。我们的安全目标是实现诚实但好奇模型下第三级别的隐私保护。

4 系统模型及基础理论

考虑由大量低耦合的传感器节点（如伯克利的 MICA 系列）自组织而成的静态网络。因此，节点资源足以存放数字节的隐私保护信息，且足以进行简单的计算操作如散列运算。节点间松散同步[12]。节点功能相似，所有节点均可能担当簇头、路由节点及处理节点。虽然传感数据具有多种类别，由于整型数据在存储和传输上更为有效，且非整型数据可以转换为整型数据，同文献[4,14]，假设传感数据为在 0 和上界 d_m 间变化的整型数据。此外，以簇为数据处理的基本单位。

4.1 系统模型

为了低能耗地实现 3.2 节中的目标，设计系统模型如下。为便于描述，首先给出相关定义。

定义 1 （前驱节点）对于距离基站 i ($i\geqslant 1$)跳的节点 b，其前驱节点为其邻居节点，且距离基站($i-1$)跳。

在所提出的机制中，节点隐藏其传感数据实现隐私保护，并通过节点间协作匿名还原出隐藏后传感数据，给出数据还原节点定义如下。

定义 2 （数据还原组 CG_i）对于任意簇，称从其簇成员的公共前驱节点中随机选取的部分节点所构成集合为数据还原组 CG_1。称从 CG_1 中成员其前驱节点的并集中随机选取的部分节点所构成集合为数据还原组 CG_2；类似地，有数据还原组 $CG_3\cdots CG_s$。

定义 3 （数据还原节点 G_i）对于任意簇，定义其 s 个共同匿名还原其隐藏后数据的节点为其数据还原节点，简称为还原节点，记为 $\{G_i(1\leqslant i\leqslant s)\}$。

对于任意簇，其还原节点 $G_i(1\leqslant i\leqslant s)$ 不是由某一个节点固定担当的，而是根据路由路径，动态地从其还原组 $CG_i(1\leqslant i\leqslant s)$ 中选取位于路由路径上的某个节点担当。在数据汇报阶段过程：簇节点隐藏所采集的数据后发送给当前还原节点 G_1；G_1 执行数据变化操作，并以匿名方式将中间值转发给当前的还原节点 G_2，如此往复直至 G_s，最终由 G_s 还原出所有数据，再进行数据处理。本文中的簇及其数据还原组易于形成，因为传感网具有树型或环型等多种路由拓扑[17,18]，所以，无论形成何种路由拓扑，节点易知其距离基站的跳数及往基站方向 s 跳以内节点和相应跳数。

4.2 隐私向量的构造和性质

基于所构造的隐私向量实现传感数据的隐藏保护和还原，接下来介绍其构造和使用。

定义 4 （隐私向量）任意节点 b 的隐私向量 $\boldsymbol{R}^b=(r^b,r^{b_1},\cdots,r^{b_s})$ 是一个(s+1)维向量，且满足 $(r^b+\sum_{i=1}^{s}r^{b_i})\bmod d_m=0(s\geqslant 2)$，其中，$d_m$ 为传感数据的上界。

任意节点 b 的隐私向量 $\boldsymbol{R}^b$ 具有机密性，具体而言，$\boldsymbol{R}^b$ 中的分量 r^b 仅为 b 所知晓；任意分量 $r^{b_i}(1\leqslant i\leqslant s)$ 则仅与 G_i 共享。b 利用 $\boldsymbol{R}^b$ 中分量 r^b 将其传感数据 d^b 隐藏为 $\widehat{d}^b=(d^b+r^b)\bmod d_m$ 以实现隐私保护；类似地，通过 $\boldsymbol{R}^b$ 中的其他分量 $(r^{b_1},\cdots,r^{b_s})$

实现隐藏数据的还原。

接下来给出的性质保证了隐私向量在传感数据的隐藏保护和还原方面的有效性。

性质 1　对于任意 d_b，记 $\hat{d}^b=(d^b+r^b)\bmod d_m$、$\hat{d}^{b_1}=(\hat{d}^b+r^{b_1})\bmod d_m$，并记 $\hat{d}^{b_j}=(\hat{d}^{b_{(j-1)}}+r^{b_j})\bmod d_m\ (2\leqslant j\leqslant s)$，则有 $\hat{d}^{b_s}=d_b$。

4.3　种子分发算法

本节给出种子分发算法的理论基础。其中，性质 2 是取模运算的代数特性；根据性质 2 易于获得性质 3，性质 3 保证了种子切分和还原方面的有效性。

性质 2　$(t_1+t_2)\bmod p=(t_1\bmod p+t_2\bmod p)\cdot\bmod p$。

性质 3　若将 r^{b_i} 切分为 s 个分量 $(r_1^{b_i},r_2^{b_i},\cdots,r_s^{b_i})$，切分满足 $(\sum_{j=1}^{s}r_j^{b_i})\bmod d_m=r^{b_i}$，并进一步将 $r_j^{b_i}(1\leqslant j\leqslant s)$ 切分为 s' 个分量 $(r_{j,1}^{b_i},r_{j,2}^{b_i},\cdots,r_{j,s'}^{b_i})$，切分满足 $(\sum_{k=1}^{s'}r_{j,k}^{b_i})\bmod d_m=r_j^{b_i}$，则 $\sum_{k=1}^{s'}((\sum_{j=1}^{s}r_{j,k}^{b_i})\bmod d_m)\bmod d_m=r^{b_i}$。

接下来以节点 b 分发仅与数据还原组 CG_i 中节点 $g_{i,w}$ 间共享的种子 $S^b\_g_{iw}$（即 $S^{b_i}\_g_{iw}$）给出种子分发算法，该算法实现了种子的机密性。本节用到的符号及含义如下。

$K_{a,b}$：节点 a 与节点 b 间共享的对偶密钥。

CG_i：ID 为 i 的数据还原组。

s_i：CG_i 的成员数目，即 $|CG_i|$。

$g_{i,w}$：CG_i 中 ID 为 w 的成员节点，下标 i 代表处理组 ID，w 代表组内 ID。

b_i：节点 b 在 CG_i 中的匿名 ID。

$S^{b_i}\_g_{iw}$：为 b 分发给 $g_{i,w}$ 的种子 $S^b\_g_{iw}$。

$s_{i,j}^{(i-1),k}\_g_{iw}$：目标节点为 $g_{i,w}$ 的种子的中间份额，由 $g_{(i-1),k}$ 生成并发送给 $g_{i,j}$；（注：$s_{i,j}^{(i-1),k}$ 上标为种子份额的生成节点标识，下标为下一跳接收节点标识）。

$\{M\}_K$：用密钥 K 加密后的消息 M。

消息分组 $\{ID_1,\{(Data,ID_2)\}_K\}$：$ID_1$ 为 Data 所属的目标节点 ID；ID_2 为 Data 当前的匿名数据 ID。

Case1　（数据源节点 b）。

Step1　切分种子 $S^b\_g_{iw}$ 为：$s_{1,1}^b\_g_{i,w}\cdots s_{1,s_1}^b\_g_{i,w}$，切分满足：$(\sum_{k=1}^{s_1}s_{1,k}^b\_g_{i,w})\bmod d_m=g_{i,w}\_s^b$。

Step2　向节点 $g_{1,k}\ (k=1,\cdots,s_1)$ 发送 $\{g_{i,w},\{(s_{1,k}^b\_g_{i,w},b)\}_{K_{b,g_{1,k}}}\}$。

Case 2　（CG_1 中任意节点 $g_{1,k}(1\leqslant k\leqslant s_1)$）。

Step1　解密 $\{b,g_{i,w},\{(s_{1,k}^b\_g_{i,w},b)\}_{K_{b,g_{1,k}}}\}$，在获得 $s_{1,k}^b\_g_{i,w}$ 后，类似 Case 1 中 b，将种子切片 $s_{1,k}^b\_g_{i,w}$ 进一步切分为：$s_{2,1}^{1,k}\_g_{i,w},\cdots,s_{2,s_2}^{1,k}\_g_{i,w}$。

Step2　向节点 $g_{2,e}\ (1\leqslant e\leqslant s_1)$ 发送 $\{g_{i,w},\{(s_{2,e}^{1,k}\_g_{i,w},b_1)\}_{K_{g_{1,k},g_{2,e}}}\}$。

Case3　（$CG_j(j=2,\cdots,(i-1))$ 中任意节点 $g_{j,v}(1\leqslant v\leqslant s_{(j-1)})$）。

Step1　解密收到的数据，一旦获得所有 ID 为 $b_{(j-1)}$ 的种子份额，合并这些份额为 $s^{j,v}\_g_{i,w}$

$$s^{j,v}\_g_{i,w}=(\sum_{u=1}^{s_{(j-1)}}s_{j,v}^{(j-1),u}\_g_{i,w})\bmod d_m$$

Step2　如果 $j=(i-1)$，即 $g_{j,v}\in CG_{(i-1)}$，则转 Step 3；否则，转 Step 4。

Step3　向 $g_{i,w}$ 发送 $\{g_{i,w},\{(s^{j,v}\_g_{i,w},b_j)\}_{K_{g_{j,v},g_{i,w}}}\}$。结束。

Step4　同 $g_{1,k}$，将 $s^{j,v}\_g_{i,w}$ 切分为：$s_{(j+1),1}^{j,v}\_g_{i,w},\cdots,s_{(j+1),s_j}^{j,v}\_g_{i,w}$，并向相应节点 $g_{(j+1),r}$ 发送 $\{g_{i,w},\{(s_{(j+1),r}^{j,v}\_g_{i,w})\}_{K_{g_{j,v},g_{(j+1),r}}}\}$。结束。

Case4　（目标节点 $g_{i,w}$）。

解密收到的数据，一旦获得所有 ID 为 $b_{(i-1)}$ 的种子份额，计算得 $S^{b_{(i-1)}}\_g_{i,w}$：$S^{b_{(i-1)}}\_g_{i,w}=(\sum_{v=1}^{s_{(i-1)}}s^{(i-1),v}\_g_{i,w})\bmod d_m$。$g_{i,w}$ 利用 $S^{b_{(i-1)}}\_g_{i,w}$ 处理 ID 为 $b_{(i-1)}$ 的数据。

$b_{(i-1)}$ 实质为节点 b 在 $CG_{(i-1)}$ 中的匿名 ID，而由性质 3 可知，$S^{b_{(i-1)}}\_g_{i,w}=S^b\_g_{iw}$。因此，$g_{i,w}$ 所存储的种子 $S^{b_{(i-1)}}\_g_{i,w}$ 实质为节点 b 分发的种子 $S^b\_g_{iw}$。

5　数据处理中的隐私保护机制 PDPV

本节具体描述提出的机制 PDPV。PDPV 基于所构造的隐私向量实现传感数据的隐藏保护和匿名的数据还原，包括分发种子的初始化和数据汇报 2 部分。

1) 初始化（详见 5.1 节）：节点执行种子分发算法，向其处理组节点分发秘密种子。仅当网络部

署或节点新加入网络时，进入初始化状态。

2) 初始化过程后，网络周期性汇报数据（详见5.2节）：各簇首先根据路由路径，动态地从其还原组中选取还原节点。接下来，节点生成隐私向量隐藏传感数据，并通过还原节点协作还原传感数据。最后，执行数据处理操作。

5.1 初始化

初始化目的是节点成簇及秘密种子分发。为实现4.1节中的模型结构，可以有多种成簇方式，在此，提供一种基于与跳数的成簇方式。本文的主要目的在于解决数据的隐私保护问题，因此，不在此对成簇的方法做过多探讨和比较。节点部署后，首先根据基站广播获取与基站间跳数，并获取往基站方向且相距跳数不超过 s 的节点信息，即节点ID及相距跳数 j $(1\leqslant j\leqslant s)$。称与基站跳数相同的节点为同级节点，初始化过程如下。

Step1 节点向其邻居节点广播其级数（即距离基站跳数）及前驱节点ID，一旦获取了所有同级邻居节点的广播信息，则计算与之前驱节点集合的交集，并计算邻居节点间前驱节点的交集，若与之相比，其交集具有最多节点，则该节点自荐为簇头。剩余节点将自荐为簇头的同级邻居节点作为候选簇头，获取与其前驱节点集合的交集，并选最大交集对应的候选簇头作为簇头，加入其簇完成成簇。成簇完成后，若簇节点数目少于 $s_{\min}$，或有节点没有成簇，则这些节点加入具有最多邻居节点的簇。

Step2 成簇后，各簇的簇节点从其公共前驱节点集合中选取 s_1 个节点作为数据处理组 CG_1，选取原则为使各簇节点发往 CG_1 的数据跳数之和最少；CG_1 从成员节点的前驱节点的并集中选取 s_2 个节点构成 CG_2，选取原则为保证各成员节点至少有一个前驱节点属于 CG_2；同法，CG_{i-1} 从其成员节点的一跳前驱节点的并集中选出 s_i 个节点构成 $CG_i(2\leqslant i\leqslant s)$。

此外，簇头随机指定 CG_1 中节点 a 为ID变化节点。节点 a 将簇节点ID变化为匿名的数据ID，并将ID变化关系加密分发给 CG_1 中其他节点；节点 a 仅将变化后的数据ID发送给 CG_2 中节点，并随机指定 CG_2 中节点 v 作为数据ID变化节点。类似地，节点 v 变化 CG_1 发送的数据ID，并将ID变化关系加密发送给 CG_2 中其他节点，仅将变化后ID发送给 CG_3。如此往复，直至 $CG_{(s-1)}$。由于任意处理组成员被俘获均会导致该组的ID变换关系被敌方获取（第6节基于此做了安全性分析），该发送方式并不会降低系统的安全性。

Step3 各节点执行种子分发算法（详见4.3节），向其数据还原组中节点分发其秘密种子。

5.2 数据汇报过程

5.2.1 数据汇报过程的形式化描述

以节点 b 给出这一过程的详细描述，并首先给出本节要用到的符号解释如下。

G_j $(1\leqslant j\leqslant s)$：$CG_j$ 中当前担当数据还原的节点，与节点 b 相距 j 跳（注：CG_j 中节点 $g_{j,w}(1\leqslant w\leqslant |CG_j|)$ 均可能担当 G_j）。

$seed_j^b$：节点 b 与还原节点 G_j 间共享的种子（注：若 G_j 由 g_{jw} 担当，那么 seed_j^b 即为种子分发算法中所分发的 $S^b\_g_{jw}$）。

$H(\cdot)$：单向散列函数。

d^b：节点 b 所采集的传感数据。

r^b：节点 b 的隐私向量的分量，用来隐藏其传感数据 d^b。

$\hat{d}^b$：节点 b 利用 r^b 隐藏其传感数据 d^b 所获得的数据。

b_j：节点 b 在 CG_j 中的匿名ID。

$\hat{d}^{b_j}$：G_j 对ID为 b_{j-1} 的数据实施还原后所得数据。

Step1 节点 b 计算 r^b（根据与各个还原节点间共享的秘密种子）。

$$r^b=d_m-(\sum_{j=1}^{s}H(seed_j^b,t))\bmod d_m$$

接下来，利用 r^b 隐藏其传感数据以实现隐私保护：$\hat{d}^b=(d^b+r^b)\bmod d_m$；最后，将 $\{(\hat{d}^b,b)\}$ 及其他需要发送的数据一起发送给 G_1。

Step2 数据还原节点 G_1 在收到 $\hat{d}^b$ 后执行还原操作：$\hat{d}^{b_1}=(\hat{d}^b+r^{b_1})\bmod d_m=(\hat{d}^b+H(seed_1^b,t))\bmod d_m$。$G_1$ 对收到的其他数据执行类似对 $\hat{d}^b$ 进行的计算，并将结果连同 $\{(\hat{d}^{b_1},b_1)\}$ 发送给 G_2。

类似地，$G_j(2\leqslant j\leqslant(s-1))$ 在收到 $\hat{d}^{b_{(j-1)}}$ 后，计算 $\hat{d}^{b_j}=(\hat{d}^{b_{(j-1)}}+r^{b_j})\bmod d_m=(\hat{d}^{b_{(j-1)}}+H(seed_j^b,t))\bmod d_m$，并在处理了所收到的所有数据后，连同 $\{(\hat{d}^{b_j},b_j)\}$ 发送给 G_{j+1}。

Step3 G_s 在收到 $\hat{d}^{b_{(s-1)}}$ 后，根据数据ID $b_{(s-1)}$ 进行数据还原：$\hat{d}^{b_s}=(\hat{d}^{b_{(s-1)}}+r^{b_s})\bmod d_m=(\hat{d}^{b_{(s-1)}}+$

$H(seed_s^b,t))\bmod d_m$。

$\hat{d}^{b_s}$ 即为最终还原得的传感数据。当 G_s 还原了所有的传感数据后，即可进行求最值及压缩等各类处理操作。

5.2.2 数据汇报过程实例

设簇节点 1、2、3 在阶段 t=1 所采集的数据分别为 137、516 及 338；d_m=1 023；G_1 将 ID[1,2,3]依次变化为[7,5,1]，G_2 将 ID[7,5,1] 依次变化为[9,4,8]。设 $H(seed_1^1,1)\bmod 1\,023=158$、$H(seed_2^1,1)\bmod 1\,023=763$、$H(seed_3^1,1)\bmod 1\,023=897$。数据汇报过程中的数据发送情况如图 1 所示，该图中大括号内数据为节点所发送的消息分组；小括号内为数据及其相应的数据 ID。以下以节点 1 为例给出数据隐藏和还原的具体过程。

Step1 节点 1 计算得 $r^1=1\,023-(158+763+897)\bmod 1\,023=228$、$\hat{d}^1=(d^1+r^1)\bmod d_m=(137+228)\bmod 1\,023=\boxed{365}$；接下来，将($\boxed{365}$, 1)发送给节点 G_1。

Step2 G_1 对($\boxed{365}$, 1)执行还原处理如下（注：G_1 将 ID 号 1 变化为 7）：$\hat{d}^7=(\hat{d}^1+H(seed_1^1,1)\bmod d_m)\bmod d_m=(365+158)\bmod 1\,023=\boxed{523}$。接下来，$G_1$ 将($\boxed{523}$,7)连同其他数据发送给 G_2。

Step3 G_2 对($\boxed{523}$,7)进行如下还原处理（注：G_2 将 ID 号 7 变化为 9）：$\hat{d}^9=(\hat{d}^7+H(seed_2^1,1)\bmod d_m)\bmod d_m=(523+763)\bmod 1\,023=\boxed{263}$。接下来，$G_2$ 将($\boxed{263}$, 9) 连同其他数据发送给 G_3。

Step4 G_3 据 ($\boxed{263}$,9) 得：$d^9=(\hat{d}^9+H(seed_3^1,1)\bmod d_m)\bmod d_m=(263+897)\bmod 1\,023=\boxed{137}$。

可见，$\hat{d}^9=d^1=\boxed{137}$，传感数据在数据处理节点 G_3 处实现了还原。而 G_3 仅知道数据 137 的匿名 ID 号 9，不能将该数据对应到其源节点 ID，不能获取节点的隐私数据。

6 安全性分析

本节分析分布式机制 PDPV 的安全性，并与 CPDA、SMART 及 KIPDA 进行比较。

6.1 PDPV 的安全性

依次分析初始安全性化阶段及数据汇报过程的安全属性，并在此基础上给出与安全属性相关的系统参数的推荐值。

6.1.1 *初始化阶段种子的安全性分析*

在初始化阶段，任意节点 b 分发给 CG_1 中任意节点 a 的种子是通过对偶密钥加密后，直接发送给节点 a 的。因此，敌方只有俘获了节点 a 才能获取该种子。对于节点 b 分发给其他数据处理组（即 $\{CG_i(2\leqslant i\leqslant s)\}$）中节点的种子，根据以下给出的定理 1，只要 $s_j+(j-1)\geqslant s(1\leqslant j\leqslant s-1)$ 成立，则敌方至少需俘获 s 个节点才能以一定概率获取该种子。

定理 1 对于任意节点 b，敌方要想获取其与 $CG_i(2\leqslant i\leqslant s)$ 中节点间所共享的种子，至少需俘获 $s_j+(j-1)$ 个节点，其中 $1\leqslant j\leqslant(i-1)$。

证明 对于任意节点 b，其与 $CG_i(2\leqslant i\leqslant s)$ 中任意节点 a 间共享的种子是在初始化阶段，通过种子切分和匿名传输实现的。因此，要想获取节点 b 与节点 a 间共享的种子，一方面，必须：1）俘获 CG_1 中的 s_1 个节点，或者 2）俘获 $CG_j(2\leqslant j\leqslant i-1)$ 中的 s_j 个节点，以还原该种子；另一方面，对于 $2\leqslant j\leqslant i-1$，俘获各个 CG_j 中至少 1 个节点以获取匿名 ID 变换关系。因此，敌方必须至少俘获 $s_j+(j-1)$ 个节点才能获取节点 b 与节点 a 间共享的种子。

由定理 1 可知，若 s 取定，即可确定 $s_i(1\leqslant i\leqslant s)$ 的最小值。例如，取定 s 为 3，则只要

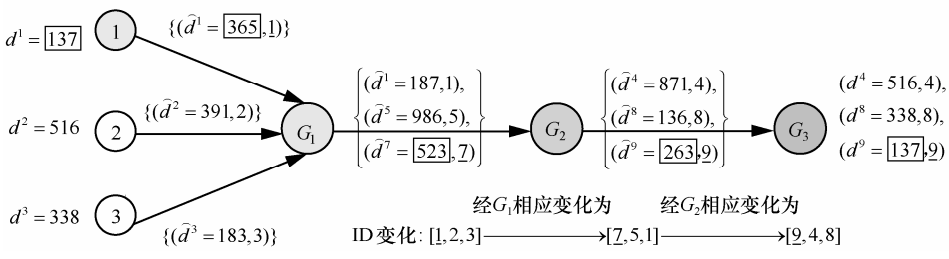


图 1　数据发送

（注：G_i 为 CG_i 中当前担当数据还原的节点；带方框的数据与节点 1 相关，其数据 ID 则加以下划线，依次为 1、7 和 9）

s_1 取 3、s_2 取 2、s_3 取 1 即可使敌方至少需要俘获 3 个节点才可能获取节点 b 的种子。在实际应用中，节点可能出现故障等意外，需保持一定的冗余性。因此，将各数据处理组所包含的节点数设为 $u(u \geqslant s)$。

6.1.2 *数据汇报阶段的隐私保护有效性*

在数据汇报阶段，传感数据被节点隐藏后匿名传输，并经多个节点协作，最终在处理节点处得到匿名还原和处理。因此，数据隐私保护的安全性取决于隐私向量的安全性及匿名 ID 的安全性。网络中节点均可能遭到俘获攻击，同已有工作[14]，不妨假设任何节点被俘获的概率相等。由接下来给出的定理 2 可知，在数据采集阶段，新提出机制 PDPV 可以有效抵御节点俘获攻击，具有良好的隐私保护属性。

定理 2 1)对于任意节点 b，敌方必须俘获至少 s 个节点才能获取 d^b。2) 若网络中被俘获节点数不少于 s，则 d^b 被敌方获取的概率 P_V 为

$$\frac{q^s(1-q^{(N-s-1)})u^{s-1}}{\mathrm{C}_N^1\cdots\mathrm{C}_{N-s+1}^1(1-q)}+q^{N-1}$$

证明 以数据的传输过程进行证明。

1) ①节点 G_1 所能获取的数据为隐藏后的传感数据 $\widehat{d}^b$（$\widehat{d}^b=(d^b+r^b) \bmod d_m$）。由于 G_1 不能获取 r^b，必不能获取 d^b。②对于任意 $G_i(2 \leqslant i \leqslant s)$，由于其只能获取隐藏后数据 $\widehat{d}^{b_i}$，且不能获取 b_i 与 b 之间的对应关系，必不能获取 d^b。要想获取 d^b，敌方必须①俘获 $\{G_i(1 \leqslant i \leqslant s)\}$，即俘获 s 个数据还原节点；或者②俘获 G_s，且俘获(s−1)个数据处理组 $\{G_i(1 \leqslant i \leqslant s-1)\}$ 中至少各 1 个节点以获取匿名 ID 变化关系。因此，敌方必须至少俘获 s 个节点才能获取 d^b。

2) 由 1)可知，若网络中被俘获节点数少于 s，敌方必不能获取 d^b。数据处理组包含多个节点，因此，1) 中情形②发生的概率高于①，于是，节点的隐私数据被敌方获取的概率 $P_V=\dfrac{\sum_{i=1}^{s-1}s_i}{\mathrm{C}_N^1\cdots\mathrm{C}_{N-s+1}^1}\sum_{i=s}^{N-2}q^i+q^{N-1}=\dfrac{q^s(1-q^{(N-s-1)})u^{s-1}}{\mathrm{C}_N^1\cdots\mathrm{C}_{N-s+1}^1(1-q)}+q^{N-1}$。

从定理 2 可知，P_V 与系统参数 s、N、q 及 u 均相关。接下来通过分析各参数对 P_V 的影响，给出各参数的推荐值。

1) 为研究处理组个数 s 对 P_V 的影响，取网络节点数 N=1 000，节点被俘获概率 q=0.1，表 1 给出了 s 在 2~7 间变化，u 在 3~7 间变化时 P_V 的变化情况。

从表 1 可知，P_V 随 s 的增长而迅速降低：s 每增长 1，P_V 则下降至少 3 个数量级。根据表 1 数据，推荐参数 s 取值为 3。此外，从表 1 可知，若 s 取定，u 增加时，P_V 随之变化的并不敏感。也就是说，适量增大 u 的取值对隐私保护有效性影响不大。因此，在节点易于出现故障的场合，可以适当增加冗余节点。

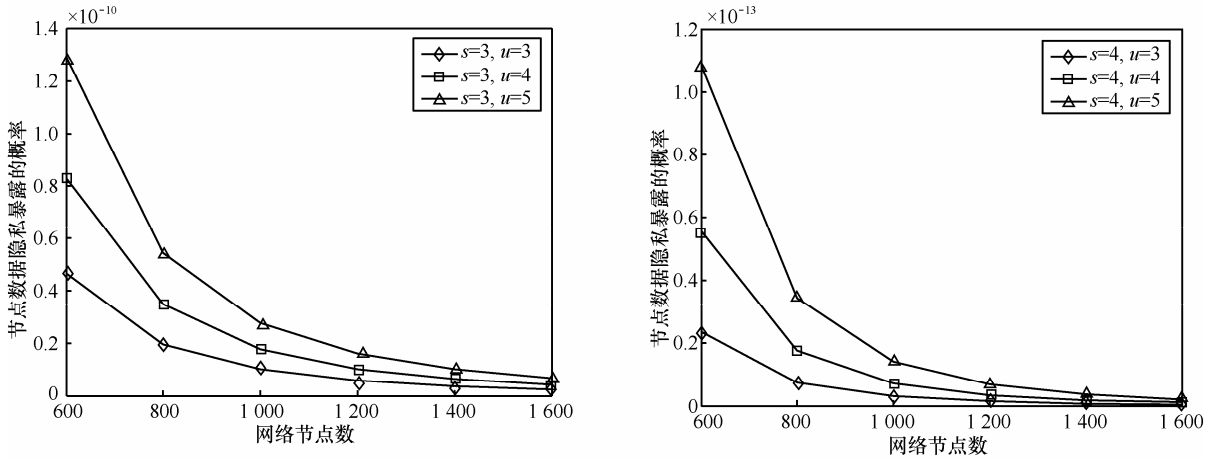
2) 为研究网络节点数 N 对 P_V 的影响，取 q=0.1，如图 2 所示，给出了 s 分别取值 3 和 4，u 取值 3~5 时，P_V 随 N 变化的的情况。从图 2 可以看出，若 s 和 u 确定，P_V 随 N 的增长而降低。这是由于网络节点数越多，敌方俘获某个节点，而该节点恰好是节点 b 的数据处理节点的概率越小。若 s 和 N 确定，则 u 越大，P_V 越低，这是由于 u 越大意味着被俘获节点恰为 b 的数据处理组中节点的概率越大。

此外，通过比较图 2(a)和图 2(b)同样可以看出，若 N 及 u 取值相同，则 s 越高，P_V 越低。

3) 为研究节点被俘获概率 q 对 P_V 的影响，取 N=1 000，图 3 给出了 s 分别取值 3 和 4，u 取值 3~5 时，P_V 随 q 的变化情况。从图 3 可以看出，若 s 和 u 取值确定，P_V 随 q 的增长而增长。若网络部署于节点易于被俘获的场合，结合表 1 可知，所提出的 PDPV 通过增大 s 的取值可有效提高隐私保护有效性。

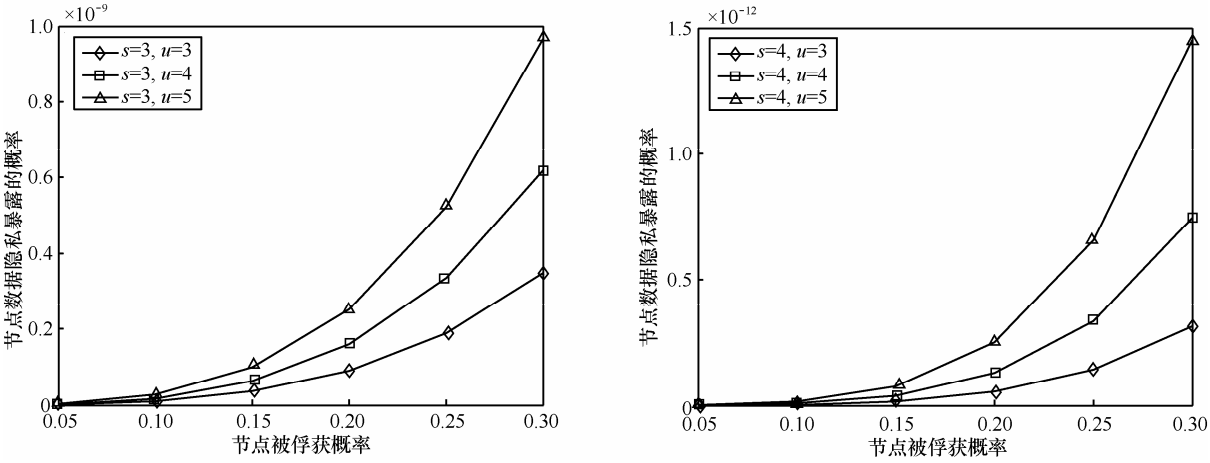
表 1 N=1 000，q=0.1，s 及 u 变化 P_V 的值（注：N 为节点数；q 为节点被俘获概率；s 为处理组个数；u 为处理组大小）

u 值	s=2	s=3	s =4	s =5	s =6	s =7
3	$3.336\,7\times10^{-8}$	$1.003\,0\times10^{-11}$	$3.018\,1\times10^{-15}$	$9.090\,6\times10^{-19}$	$2.740\,9\times10^{-22}$	$8.272\,3\times10^{-26}$
4	$4.448\,9\times10^{-8}$	$1.783\,1\times10^{-11}$	$7.154\,0\times10^{-15}$	$2.873\,1\times10^{-18}$	$1.155\,0\times10^{-21}$	$4.647\,9\times10^{-25}$
5	$5.561\,1\times10^{-8}$	$2.786\,1\times10^{-11}$	$1.397\,3\times10^{-14}$	$7.014\,3\times10^{-18}$	$3.524\,8\times10^{-21}$	$1.773\,0\times10^{-24}$
6	$6.673\,3\times10^{-8}$	$4.012\,0\times10^{-11}$	$2.414\,5\times10^{-14}$	$1.454\,5\times10^{-17}$	$8.770\,8\times10^{-21}$	$5.294\,3\times10^{-24}$
7	$7.785\,6\times10^{-8}$	$5.460\,8\times10^{-11}$	$3.834\,1\times10^{-14}$	$2.694\,6\times10^{-17}$	$1.895\,7\times10^{-20}$	$1.335\,0\times10^{-23}$



(a) 处理组数目 s=3；处理组大小 u=3~5　　(b) 处理组数目 s=4；处理组大小 u=3~5

图 2　节点数据隐私暴露的概率随网络节点数 N 的变化关系



(a) 处理组数目 s=3；处理组大小 u=3~5　　(b) 处理组数目 s=4；处理组大小 u=3~5

图 3　数据隐私暴露的概率随节点被俘获概率 q 的变化关系

6.2　隐私保护有效性比较

接下来分析 CPDA、SMART 及 KIPDA 的隐私保护有效性，并与之进行比较。

机制CPDA通过簇内节点间协作实现数据隐私保护。具体而言，在数据汇报阶段，簇节点将数据隐藏后加密发送给各个簇成员；而一旦收到其他所有簇节点发送的数据，则进行求和操作并将结果发送给簇头，由簇头还原出和值。对任意节点 b 而言，若其所在簇的其他成员均被俘获，那么其数据隐私将会暴露。记 CPDA 的簇大小为 n_c，则其所能容忍的被俘获节点数为 (n_c-1)，其节点隐私数据暴露的概率 P_C 为 $\sum_{j=n_c-1}^{N-1}\frac{q^j \mathrm{C}_j^{n_c}}{\mathrm{C}_N^j}$。

为实现数据隐私保护，SMART 中节点将其传感数据切分为 J 块，并将其中的$(J-1)$块加密后分别发送给$(J-1)$个邻居节点。可见，若任意节点 b 的入度与出度节点均被俘获，则其数据隐私将会暴露。因此，SMART 所能容忍的被俘获簇节点的平均数目 J_s 为 $\frac{3(J-1)}{2}$，其节点数据隐私暴露的平均概率 P_S 为 $\sum_{j=J_s-1}^{N-1}\frac{q^j \mathrm{C}_j^{J_s}}{\mathrm{C}_N^j}$。

机制 CPDA 及 SMART 其隐私保护有效性同所提出机制 PDPV 一样，仅受被俘获节点数影响。若 $n_c \leqslant s$，那么 PDPV 比 CPDA 可以容忍更多的被俘获节点，即有着更强的隐私保护力度。若 $J \leqslant s$，则 PDPV 比 SMART 有着更强的隐私保护力度。而值得一提的是，CPDA 及 SMART 仅适用于求和处理。

机制 KIPDA 的隐私保护有效性受伪装度 k、消息分组大小 $|I|$ 及被俘获节点数目 c 等多个参数的影响，较为复杂。机制 KIPDA 通过 k-伪装实现传感

数据的隐私保护：若敌方俘获了某个节点，则通过窃听可获知其他节点的传感数据是消息分组中特定的 k 条之一，其中，消息分组包含 $|I|$ 条数据，且 k 远小于 $|I|$。可见，k 越高，则伪装保护越有效。而若敌方俘获了多达 c 个节点，则可以获知所有节点的隐私数据，可知，任意节点的隐私数据被破获的概率 $P_K=\sum_{i=c}^{N-1}q^i=\frac{q^c(1-q^{(N-c)})}{1-q}$。$P_K$ 随 c 的增长而增长；然而，当 $|I|$ 固定时，参数 c 与 k 是一对矛盾，若增加 k，则 c 将会降低（详见文献[14]）。KIPDA 的通信能耗与 $|I|$ 正相关，因此，文献[14]中将 $|I|$ 的值推荐为 15，并对 $|I|$ 固定为 15 时参数 c 随 k 的变化情况进行了研究，结果显示：1) k=3 时，c=11；2) k=4 时，c=8；3) k=5 时，c=6。

与 KIPDA 相比，机制 PDPV 对任意包含 n_v 个节点的簇实现了强度更高的 n_v-伪装保护，且只有敌方俘获了其处理节点（非任意节点），才能获知其簇节点的传感数据是 n_v 个数据中的某一个。这是因为任何路由节点不能获取传感数据；且即使是处理节点也仅能获知特定簇节点的数据为其所还原的 n_v 个数据之一。不难得出结论，若 $n_v>k$，则 PDPV 具有更好的伪装保护性能。此外，对于 PDPV，只有不少于 s 个节点被敌方俘获后，敌方才能以概率 P_K 获取节点的隐私数据。接下来给出的定理 3 对 P_V 和 P_K 在理论上进行了比较。

定理 3 若 $s\geqslant c$，则有 $P_K>>P_V$（>>代表远大于）；即使 $c>s$，若 $\left(\frac{N-s+1}{u}\right)^{s-1}>q^{(s-c)}$ 成立，则 $P_K>P_V$ 仍然成立。

证明 对于 P_V 和 P_K，有

$$\frac{P_K}{P_V}>\frac{P_K-q^{N-1}}{P_V-q^{N-1}}$$

$$=\frac{C_N^1\cdots C_{N-s+1}^1q^{(c-s)}(1-q^{(N-c-1)})}{u^{(s-1)}(1-q^{(N-s-1)})}$$

$$>\left(\frac{N-s+1}{u}\right)^{s-1}\frac{q^{(c-s)}(1-q^{(N-c-1)})}{(1-q^{(N-s-1)})}$$

1) 若 $s\geqslant c$，则 $\frac{(1-q^{(N-c-1)})}{(1-q^{(N-s-1)})}>1$，且 $q^{(c-s)}\geqslant 1$；此外，$N-s+1$ 是远大于 u 的，则有 $\left(\frac{(N-s+1)}{u}\right)^{s-1}>>1$，于是 $P_K>>P_V$。

2) KIPDA 中的系统参数 c 通常是一个较小的值，这是因为 c 受限于通信能耗及伪装力度 k，不能随意增长；而 N 通常远远大于 s 和 c，且 $q<1$，则有 $\frac{(1-q^{(N-c-1)})}{(1-q^{(N-s-1)})}$ 趋近于 1。于是，若 $\left(\frac{(N-s+1)}{u}\right)^{s-1}>q^{(s-c)}$ 成立，则有 $P_K>P_V$。

可见，若 $s\geqslant c$，则 PDPV 在对抗节点俘获攻击方面要远远优于 KIPDA；而对于 $c>s$ 的情况，如算例 1 所示，在典型的参数设置下，P_V 同样低于 P_K。

算例 1 取 N=1 000，$q=0.1$，1）对于 CPDA 和 SMART，取其推荐值即簇内节点数 n_c=3，数据分片数 J=3 进行计算，可得两者的数据隐私暴露概率分别为 P_C=2.003 8×10^{-8} 和 P_S=2.003 8×10^{-8}。2）对于 KIPDA，取其推荐值进行计算：$|I|$ =15，①k=3，c =11，则 P_K=1.11×10^{-11}；②k =4，c =8，则 P_K=1.11×10^{-8}；③k=5，c=6，则 P_K=1.11×10^{-6}。3）对于 PDPV，①取 s=3，u=4，有 P_V=1.783 1×10^{-11}；②取 s=3，u=6，有 P_V=4.012×10^{-11}。

由算例 1 可知，在典型的参数取值下，与 KIPDA 相比，即使其参数 k 取 3，PDPV 仍然具有更好的抵御节点俘获攻击性能；而簇大小 n_v 通常于 5，这意味着 PDPV 具有比 KIPDA 更好的伪装性。可见，PDPV 的隐私保护性能均优于 KIPDA。此外，PDPV 抵御节点俘获攻击的能力同样要优于 CPDA 和 SMART。

7 系统开销

本节分析和比较各机制中单个节点所引发的系统开销。为便于描述，记传感数据长 L_{sen} bit，PDPV 及 CPDA 中的簇大小分别为 n_v 和 n_c。

7.1 通信开销

PDPV 中的节点消息分组包含数据及其数据 ID，记数据 ID 长 $l_{\text{clu}}=\lceil\log n_v\rceil$ bit，消息分组在抵达处理节点前需要经历 s 跳，于是，通信开销为 $s(L_{\text{sen}}+l_{\text{clu}})$ bit。可见，通信开销随 s 线性增长，且增长率低；而由 6 节可知，若 s 增大 1，则机制在对抗节点俘获攻击方面的隐私保护有效性可以提升数个数量级，因此，PDPV 的隐私保护有效性并未受限于通信开销。对于参数 n_v，通信开销随之变化的并不明显：例如，1）取 s=3，n_v =16，则通信

开销为 $3[L_{\text{sen}}+4]$ bit；2）取 s=3，n_v =28，通信开销也仅为 $3[L_{\text{sen}}+5]$ bit。因此，将 n_v 推荐为 12~32。在伪装保护力度方面，PDPV 在 n_v 取推荐值时比 KIPDA 取推荐值 k 时更为有效。

KIPDA 中的节点消息分组包含 $|I|$ 条数据，故其通信开销为 $|I|L_{\text{sen}}$ bit。在典型参数配置下，PDPV 的通信开销优于 KIPDA。这是由于：为保持有效的伪装性能，$|I|$ 须在特定值以上，通常 $|I|>s$；且数据 ID 长 l_{clu} 通常远小于传感数据长 L_{sen}。例如，文献[14]中 $|I|$ 的推荐值为 15，若 s 同样取推荐值 3，有：KIPDA 的通信开销为 $15L_{\text{sen}}$ bit；PDPV 的通信开销为 $3(L_{\text{sen}}+l_{\text{clu}})$ bit，明显优于 KIPDA。而由第 6 节可知，此时 PDPV 在抗节点俘获攻击及数据伪装保护性能方面均同样优于 KIPDA。

CPDA 中的任意节点 b 需要将隐藏后的传感数据 $\{V_c^b,b\}$ 发送给任意簇成员 c，其中，$|V_c^b|\geqslant L_{\text{sen}}$；节点 b 还需要将隐藏后的传感数据 $\{F_b,b\}$ 发送给簇头，其中，$|F_b|\geqslant(L_{\text{sen}}+l_{\text{clu}})$。于是，各节点总的通信开销为 $n_c(L_{\text{sen}}+2\lceil\log N\rceil)$ bit。若 $n_c=s$，则 PDPV 与 CPDA、SMART 的通信开销基本相当，却能容忍更多的被俘获节点。

SMART 中节点将其传感数据分为 J 块（每个块数据的长 L_{sen} bit），并将其中的(J−1)块分别发送给(J−1)个邻居节点。此外，该节点将所收到的数据及所保留的块数据进行聚合并发送给下一跳节点，该数据长为 $(L_{\text{sen}}+\lceil\log N\rceil)$ bit。因此，各节点至少需要发送 J 个消息分组，总的通信开销为 $J(L_{\text{sen}}+\lceil\log N\rceil)$ bit。若 $J=s$，则 PDPV 与 CPDA、SMART 的通信开销基本相当，却能容忍更多的被俘获节点。

7.2　存储开销

PDPV 中的节点需要存储其种子；该节点若担当数据处理候选节点，还需要存储所处理簇的种子（最多 s 个）。因而，节点的存储开销最大为 $(\sum_1^s s_i+sn_v)L_{\text{sen}}=s(u+n_v)L_{\text{sen}}$ bit。SMART 及 CPDA 中节点无需额外的存储开销。KIPDA 中各节点需要存储数据索引集合 NSS，其存储开销为 $\log\lceil|I|\rceil$ bit。可见，PDPV 的存储开销高于 SMART、CPDA 及 KIPDA，然而，其存储开销仍然是适合传感网的。例如，取 $L_{\text{sen}}=10$ bit（传感数据范围为 0~1 024）；n_v =32；$s=3$，$u=5$，易知其存储开销不超过 1 110 bit。

7.3　计算开销

本节评估单个节点在数据汇报过程中额外的计算开销。PDPV 额外的计算开销主要在于生成隐私向量及数据还原操作：节点生成其隐私向量需要 s 次散列运算及复杂度为 $o(s)$ 的四则运算；此外，节点的 s 个还原节点各需执行一次散列运算，并对隐藏后数据执行复杂度为 $o(s)$ 的四则运算。因此，各节点所引发的额外的计算开销为 $2s$ 次散列运算及复杂度为 $o(s)$ 的四则运算。

KIPDA 中各节点需要填充哑数据，哑数据包含 $\overline{NSS}$ 和 NSS，且 NSS 中数位需满足限定条件，这一过程的计算开销为 $(|NSS|-1)$ 次比较操作。此外，各节点数据需要进行 $|I|$ 次比较操作。于是，总的计算开销为 $(|NSS|-1)+|I|$ 次比较操作。

CPDA 中节点需加解密 (n_c-1) 个簇成员的数据，并进行四则运算，将最终值加密发送给簇头。簇头则需解密收到的数据并还原和值，这一过程需 n_c 次加/解密和 1 次矩阵求逆。因此，各节点所引发的额外的计算开销至少为 $(2n_c-1)$ 次加/解密及复杂度为 $o(n_c)$ 的四则运算。

SMART 中节点将其传感数据切分为 J 份并将其中的(J−1)份加密发送给邻居节点；收到其数据的节点需执行解密、求和操作，并加密和值发送给其他节点。因此，各节点所引发的额外的计算开销为 $2J$ 次加/解密及复杂度为 $o(J)$ 的四则运算。

可见，PDPV 的计算开销低于 CPDA 及 SMART，高于 KIPDA。然而，PDPV 的计算开销主要源于 $2s$ 次散列运算，仍然是轻量的，适合于传感网。

7.4　能耗

传感器节点的能量有限性使得能耗是最为敏感的评估因素，节点能耗由计算能耗和通信能耗组成。接下来依托典型传感器节点比较各机制能耗。不同的加密算法具有不同能耗，经典文献[19]推荐传感器节点采用 RC5 作为加密算法。因此，此处采用 RC5 计算相关机制能耗。PDPV 采取散列运算实现隐藏数的动态性，所采用的散列函数无需考虑碰撞性这一签名等应用领域非常关注的性能；此外，节点发送的数据为散列值与传感数据模加运算后的数据，敌方难以直接获取散列值。因此，采用轻量的散列算法如 MD2，或利用取模、乘法取整及平方取中等常见的散列函数的组合对种子进行变化一样可满足需求。而为了便于比较，此处采用广为

熟知的安全散列算法 SHA-1 计算 PDPV 能耗。在实际应用中，采用其他轻量的散列运算的能耗将远低于 SHA-1。

由 Crossbow 公司开发的 MICA2dot 是典型的传感器节点，文献[20]给出了 MICA2dot 执行改进的 SHA-1 散列运算的能耗及通信能耗；文献[14]给出了 Atmega 128L 执行 RC5 算法的能耗值，具体如表 2 所示。

记对每 1bit 数据执行单向散列运算所需能耗为 $Hash(\text{bpv})$，则 PDPV 中各节点能耗为：$2s(L_{\text{sen}}+l_t)Hash(\text{bpv})+s(L_{\text{sen}}+l_{\text{clu}})(R(\text{bpv})+T(\text{bpv}))$（$l_t$ 为阶段数长度）；忽略 KIPDA 的计算开销能耗，则 KIPDA 中各节点能耗为：$|I|L_{\text{sen}}(R(\text{bpv})+T(\text{bpv}))$。记节点加解密 1 bit 所需能耗分别为 E_{ENC} 和 E_{DEC}，则 CPDA 和 SMART 的能耗分别为 $n_c(L_{\text{sen}}+2\lceil N\rceil)(R(\text{bpv})+T(\text{bpv}))+(n_c-1)L_{\text{sen}}E_{\text{DEC}}+n_c L_{\text{sen}}E_{\text{ENC}}$；$J(L_{\text{sen}}+l_{\text{clu}})(R(\text{bpv})+T(\text{bpv}))+J L_{\text{sen}}(E_{\text{DEC}}+E_{\text{ENC}})$。虽然加密及散列算法在实际执行中能耗为固定位数如 64 bit 的倍数，为便于比较，按能耗/bit 计算 CPDA 等机制单个数据的计算能耗。这样计算是合适的，因为在实际数据汇报，特别是周期性数据汇报过程中，通常多个数据一起发送，实际数据长可为固定数位的倍数，这样单个数据的计算能耗即为此处计算的数据长与能耗/bit 的乘积。文献[14]中，$|I|$的推荐值为 15，L_{sen} 的推荐值为 10 bit；在文献[4]中，n_c 的推荐值为 3，J 的推荐值为 3；取 $l_t=16$，s=3，$l_{\text{clu}}=5$ bit，根据表 2 中值可得各机制取推荐值时的能耗如表 3 所示。

从表 3 可知，PDPV 具有比 KIPDA、CPDA 及 SMART 更低的能耗。而此时，与 KIPDA、CPDA 及 SMART 相比，PDPV 在抵御节点串谋攻击方面更强；PDPV 在伪装度方面优于 KIPDA，稍弱于 CPDA 及 SMART。

8 仿真实验

为获取整个系统执行 PDPV 所需通信开销，同已有工作，对 PDPV 在典型参数下的成簇及数据汇报过程进行仿真实验。试验基于 MATLAB 平台。如图 4 所示，总数为 1 024 的传感器节点均匀分布于 400 m×400 m 的一个监测区域，基站 BS 位于区域的左下角，每个节点的通信半径为 50 m，最小簇大小 $s_{\min}$ 取 5，s 取 3；各簇的数据处理节点在获取了处理值后，将处理值发送给下一级的数据处理节点进行进一步处理。$s_{\min}$ 的取值也可根据需要调高，这样可以增强伪装度，代价是小部分节点数据需要多一跳开销抵达还原节点 G_1。

图 4(a)显示了网络部署分级后各级节点的分布情况，图中位置相邻而级别不同的节点用不同的图形符号加以区分。例如，第一级和第二级节点分别用“□”和“×”进行标识，以示区分。从图 4(a)可以看出，网络节点共分为 13 级，每一级趋近为以 BS 为中心的扇形环。在此基础上，执行 5.1 节中的成簇算法，成簇效果如图 4(b)所示，为区分不同的簇，图中位置相邻而属于不同簇的节点用不同的图形符号加以区分。从图 4(b)可知，簇节点数目基本在 10 左右，大部分簇所覆盖区域趋近理想的

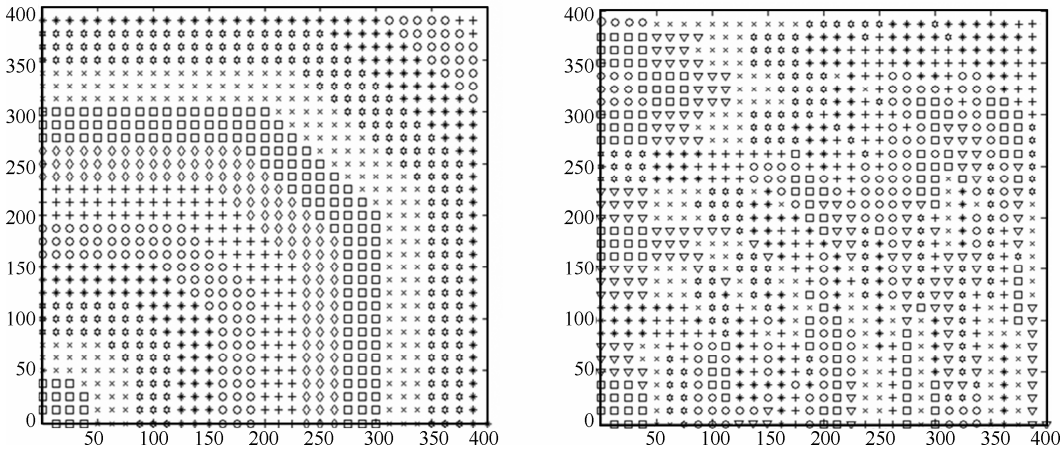
表 2 MICA2dot 的能耗（单位：μJ/bit；*T*/bpv 和 *R*/bpv 分别为发送和接收 1 bit 信息所需能耗）

T/bpv	*R*/bpv	一个时钟周期	散列运算	加密	解密
7.4	3.58	3.54×10^{-3}	0.737 5	18.15	18.14

表 3 参数取推荐值时各机制的能耗及隐私保护属性比较

机制名	非线性处理	存储/bit	能耗/μJ	伪装度 *k*	隐私暴露概率
KIPDA	适用	150	1 647	窃听：k=15 俘获任意节点：k=4（任意节点被俘获概率：0.1）	1.11×10^{-8}
PDPV	适用	1 110	604	窃听：$k=\infty$ 俘获特定处理节点：$5<k$（处理节点被俘获概率：10^{-4}）	2.79×10^{-11}
CPDA	不适用	无	1 588	窃听：$k=\infty$ 俘获任意某个节点：$k=\infty$	$2.003\ 8\times10^{-8}$
SMART	不适用	无	1 770	窃听：$k=\infty$ 俘获任意某个节点：$k=\infty$	$2.003\ 8\times10^{-8}$

注：若伪装度为 k，则敌方能猜测出节点的隐私数据为 k 个数据中的某一个值



(a) 网络节点分级示意　　(b) 网络成簇示意

图 4　网络节点分级及成簇示意

线型，这些簇节点通过一跳即可与其还原节点 G_1 通信；此外，对于区域边缘的小部分簇，其所覆盖区域大于或接近单个节点的通信范围，使部分簇节点需要通过其他簇成员中转才能将数据发送至还原节点 G_1。然而，这种节点数目较少，这意味着，只有小部分节点的通信开销大于 s 跳。虽然可以通过一些优化措施增强成簇效果，从而优化通信能耗，由于本文的主要目的在于解决数据的隐私保护问题，本实验不再对成簇算法和结果做更多的优化工作，仅直接执行 5.1 节中的成簇算法。

表 4 给出了典型参数下，系统在不同节点密度下的通信能耗。从表 4 可知，与单个节点通信能耗与节点数目的乘积相比，整个系统通信能耗的实验值略高，这与从图 4(b)直观得出的结论是一致的。此外，节点越稠密，实验值与理论值越接近，这是由于节点越稠密，节点的前驱节点越多。

表 4　　机制 PDPV 的通信能耗

节点数目	网络级数	整个网络的通信能耗（理论值）/μJ	整个网络的通信能耗（实验值）/μJ
768	12	3.79×10^5	4.22×10^5
1 024	13	5.06×10^5	5.39×10^6
280	11	6.31×10^5	6.51×10^6

9　结束语

相比于求和中的数据隐私保护，非线性处理中的隐私保护问题更为富于挑战性。本文基于所构造的隐私向量及其生成方法，提出了分布式机制 PDPV，该机制不仅适用于求和等线性处理函数，也适用于非线性处理函数。表 3 总结了 PDPV 及同类机制的各项性能，从中可知，与仅适用于求和处理的分布式机制 CPDA 及 SMART 相比，PDPV 不仅能更好地抵御节点俘获攻击且更为能量有效；与同样适用于非线性处理的机制 KIPDA 相比，PDPV 抵御节点俘获攻击性能更优，数据伪装度更好，且能耗更低。PDPV 的存储开销要高于 CPDA、SMART 及 KIPDA，然而，其存储开销是适合于传感网的。对传感网而言，能耗和安全性是最为敏感的要素，因此，PDPV 更适用于资源有限且应用相关的传感网。

参考文献：

[1] 杨庚, 许健, 陈伟, 等. 物联网安全特征与关键技术[J]. 南京邮电大学学报(自然科学版), 2010, 30(4):20-29.
YANG G, XU J, CHEN W, *et al*. Security characteristic and technology in the Internet of things[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2010, 30(4):20-29.

[2] 李仁发, 谢勇, 李蕊, 等. 信息-物理融合系统若干关键问题综述[J]. 计算机研究与发展, 2012, 49(6): 1149-1161.
LI R F, XIE Y, LI R. Survey of cyber-physical systems[J]. Journal of Computer Research and Development, 2012, 49(6): 1149-1161.

[3] 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.
FANG Y J, CHEN H, ZANG X Y. Data privacy preservation in wireless sensor networks[J]. Chinese Journal of Computers, 2012, 35(6): 1131-1146

[4] HE W, LIU X, NGUYEN H, *et al*. PDA: privacy-preserving data aggregation in wireless sensor networks[A]. Proceedings INFOCOM 2007: 26th IEEE International Conference on Computer Communications[C]. Piscataway: IEEE Press, 2006. 165-168.

[5] 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.
YANG G, WANG A Q, CHEN Z Y, *et al*. An energy_saving privacy-preserving data aggregation algorithm[J]. Chinese Journal of Computers. 2011,34(5):792-800.

[6] 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.
ZHOU Q, YANG G, LI S, *et al*. An integrity-checking private data aggregation algorithm[J]. Journal of Electronics & Information Technology, 2013,35(6):1277-1283

[7] CASTELLUCCIA C, CHAN A, MYKLETUN E, *et al*. Efficient and provably secure aggregation of encrypted data in wireless sensor networks[J]. ACM Transactions on Sensor Networks, 2009,5(3): 1-36.

[8] FENG T, WANG C, ZHANG W, *et al*. Confidentiality protection schemes for data aggregation in sensor networks[A]. Proceedings INFOCOM 2008: 27th IEEE International Conference on Computer Communications[C]. Piscataway: IEEE Press, 2008. 131-146.

[9] CONTI M, ZHANG L, ROY S, *et al*. Privacy-preserving robust data aggregation in wireless sensor networks[J]. Security and Communication Networks, 2009, 2(2): 195-213.

[10] 曾玮妮, 林亚平, 何施茗等. 无线传感器网络中基于隐私保护元的数据聚合机制[J]. 通信学报, 2012, 10(6): 16-25.
ZENG W N, LIN Y P, HE S M, *et al*. A data aggregation scheme based on privacy- preserving element for wireless sensor networks[J]. Journl on Communictions, 2012,10(6):16-25.

[11] JUNG T, MAO F, LI X, *et al*. Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation[A]. Proceedings INFOCOM 2013: 32th IEEE International Conference on Computer Communications[C]. Piscataway: IEEE Press, 2013. 2634-2642.

[12] GREUNEN J, RABAEY J. Lightweight time synchronization for sensor networks[A]. Proceedings of the Second ACM International Workshop on Wireless Sensor Networks and Applications, WSNA 2003[C]. Association for Computing Machinery, 2003. 11-19.

[13] 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, PerCom 2008[C]. Hongkong, China, 2008.

[14] GROAT M, HE W, FORREST S. KIPDA: *k*-indistinguishable privacy-preserving data aggregation in wireless sensor networks [A]. Proceedings INFOCOM 2011: 30th IEEE International Conference on Computer Communications[C]. Piscataway: IEEE Press, 2011. 2024-2032.

[15] ZHANG W, TRAN M N, ZHU S, *et al*. A random perturbation-based scheme for pairwise key establishment in sensor networks[A]. MobiHoc'07: Proceedings of the Eighth ACM International Symposium on Mobile Ad Hoc Networking and Computing[C]. New York, Association for Computing Machinery, 2007. 90-99.

[16] ALI F, MEHDI B, HOSSEIN S, *et al*. A high performance and intrinsically secure key establishment protocol for wireless sensor networks[J]. Computer Networks, 2011, 55(8): 1849-1863.

[17] MADDEN S, FRANKLIN M, HELLERSTEIN J, *et al*. Tag: a tiny aggregation service for ad-hoc sensor networks[J]. SIGOPS Oper yst Rev, 2002, 36: 131-146.

[18] GOBRIEL S, KHATTAB S, MOSSE D, *et al*. Ridesharing: fault tolerant aggregation in sensor networks using corrective actions[A]. Proceeding of the Sensor and Ad Hoc Communications and Networks, 2006. SECON '06[C]. Reston, VA, 2006.

[19] PERRIG A, SZEWCZYK R, WEN V, *et al*. SPINS: security protocols for sensor networks[A]. Proceedings of the Seventh Annual International Conference on Mobile Computing and Networks[C]. Rome, Italy, 2001.

[20] WANDER A S, GURA N, EBERLE H, *et al*. Energy analysis of public-key cryptography for wireless sensor networks[A]. Proceedings - Third IEEE International Conference on Pervasive Computing and Communications, PerCom 2005[C]. Kauai Island, HI, United States, 2005.

作者简介：

曾玮妮（1982-），女，湖南邵阳人，博士，中国船舶重工集团第 716 研究所高级工程师，主要研究方向为网络安全、可信计算。

林亚平（1955-），男，湖南邵阳人，博士，湖南大学教授、博士生导师，主要研究方向为通信网络、机器学习。

易叶青（1976-），男，湖南邵阳人，博士，湖南人文科技学院副教授，主要研究方向为传感器网络安全。

何施茗（1986-），女，湖南永州人，湖南大学博士生，主要研究方向为无线网络。

陈鹏（1977-），男，湖北十堰人，博士，中国船舶重工集团第 716 研究所高级工程师，主要研究方向为多媒体网络。