无线传感器网络多应用场景下的安全数据融合方案

陈燕俐[1,2] 张 乾[1] 许 建[1,2] 王梦涵[1]
(南京邮电大学计算机学院 南京 210003)[1]
(南京邮电大学宽带无线通信与传感网技术教育部重点实验室 南京 210003)[2]

摘 要 针对无线传感器网络多应用场景下异构数据的安全融合问题,提出了一种轻量级的安全数据融合保护方案,该方案可同时保障数据的隐私性、完整性和新鲜性。首先,以当前融合轮数和节点预置密钥作为哈希函数的输入,为节点更新每个融合周期的密钥;其次,采用同态加密技术,使中间节点能够对密文直接执行融合操作;然后,采用同态消息认证码,使基站能够验证融合数据在传输过程中是否被篡改;进一步,对明文信息采用编码机制,以满足多应用场景下异构数据聚集的使用需求。理论分析和仿真结果表明,该算法具有较好的安全性、较低的通信开销和更高的融合精确度。

关键词 安全数据融合,同态加密,同态消息认证码,多应用

中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.031

Secure Data Aggregation Scheme for Multiple Applications in Wireless Sensor Networks

CHEN Yan-li[1,2] ZHANG Qian[2] XU Jian[1,2] WANG Meng-han[1]
(College of Computer, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)[1]
(Key Lab of Broadband Wireless Communication and Sensor Network Technology,
Ministry of Education, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)[2]

Abstract To solve the issues of security of multi-source heterogeneous data during data aggregation in wireless sensor networks(WSNs), this paper proposed a lightweight secure data aggregation scheme which can guarantee data confidentiality, integrity and freshness. HASH function is used to update the key of each time slot by using present aggregation round and preset key as an input. The application of homomorphic encryption makes intermediate node perform aggregation operation on ciphertext directly. Homomorphic message authentication code (HMAC) enables base station to verify whether the aggregation data have been modified during transportation. Moreover, plaintext is coded before being encrypted so as to satisfy multiple applications. Theoretical analysis and simulation verify that the proposed algorithm can preserve data privacy with lower communication assumption and higher data aggregation accuracy.

Keywords Secure data aggregation, Homomorphic encryption, Homomorphic message authentication code(HMAC), Multiple applications

1 引言

近年来,随着无线通信与计算机技术的发展,无线传感技术也得到了大幅提高。无线传感器网络由于可自组织、易部署、节点体积微小和价格低廉等特征被广泛应用于环境监测、战场监视和目标跟踪等领域[1]。由于传感器节点在能源、计算和通信能力等方面的作用有限,各种旨在减少网络能源消耗的方法层出不穷。其中,数据融合作为WSN中的一项关键技术,大大减少了网络中的通信与计算开销。数据融合允许中间转发节点对其子节点的数据进行处理,以达到降低数据冗余、减少数据包发送量,进而延长整个网络生命周期的目的[2]。

当WSN应用于敏感数据监测时,如何保障数据的安全性成为了数据融合技术面临的主要问题之一。相对地,在保障数据安全性的同时,一般会给网络带来额外的开销,然而这与引入数据融合技术的初衷背道而驰。由于WSN的特殊

到稿日期:2016-08-09 返修日期:2016-12-04 本文受国家自然科学基金项目(61572263),江苏省高校自然科学研究项目(14KJB520031,15KJB520027),江苏省自然科学基金青年基金项目(BK20130096),南京邮电大学科研项目(NY215097,NY214127)资助。

陈燕俐(1969-),女,博士,教授,主要研究方向为网络信息安全、隐私保护,E-mail:chenyl@njupt.edu.cn;张 乾(1990-),男,硕士生,主要研究方向为信息安全、数据融合;许 建(1980-),男,博士,副教授,主要研究方向为计算机网络、信息安全、隐私保护;王梦涵(1996-),女,主要研究方向为数据融合、隐私保护。

性，一些传统的数据安全保护方案因昂贵的通信与计算开销而不适用于资源受限的无线传感器网络。因此，如何保证数据融合的安全性成为新的挑战，亟待寻求一种简单高效的方法来解决这一矛盾。另一方面，在很多监测过程中，感知系统需要采集的数据往往是复合型数据，即对多种不同类型的应用数据同时进行采集、分析与处理。在这一过程中，如何做好多源异构数据的编码、封装，并在进行数据融合处理的同时保证数据的安全是研究的热点和难点。

2　相关工作

无线传感器网络数据融合的安全需求主要包括：数据的完整性、机密性、新鲜性、可用性、准确性和不可否认性等[3]。近年来，数据安全融合研究主要致力于数据的机密性和完整性两方面。对于数据机密性的保护通常采用对数据加密的方法。然而，传统的逐跳加密方法并不能在保证数据机密性的同时执行数据融合操作，而同态加密的出现解决了这一问题。将同态加密应用于无线传感器网络，终端节点采集到隐私数据后对其直接加密后上传，融合节点可以利用同态加密的特性直接对接收到的多个密文进行融合操作，不仅保证了端到端的数据机密性，而且可有效降低网络时延[4]。目前，基于同态加密的安全数据融合已有不少研究成果。

大部分已有研究成果仅仅针对单一类型数据的安全融合问题，其中具有代表性的算法包括 CMT 算法[5]、CDA 系列算法[6-7]等。CMT 算法是由 Castelluccia 等人在文献[5]中提出的一种基于流密钥的对称同态加密方案。在该方案中，每个节点与基站共享一个密钥，节点感知的明文数据通过与基站共享的密钥进行模加法运算得到相应的密文，基站在收到融合密文后，减去所有响应节点的密钥即可得到最终融合结果。但是该方案存在以下不足：1）当只有部分节点不响应时，需要上传不响应节点的 ID，这会给网络带来额外的通信开销；2）基站无法验证数据的完整性。针对该算法存在的问题，Verma 等人提出一种优化的 ID 传输机制[8]，减少了因上传 ID 带来的额外通信开销。CDA 算法[6]则可以支持不同操作，如求平均、求方差等，但在该算法中基站节点不能获取到原始信息；Chen 等人在文献[7]中对其进行改进并提出了可恢复原始数据的隐私保护算法 RCDA。该方案使用椭圆曲线上的 ElGamal 算法（EC-EG）对明文信息进行加密，保证了数据的机密性；通过对明文进行签名，保证了数据完整性；通过对明文信息进行编码，使基站能够从融合结果中恢复出所有节点的原始数据。但该算法在计算复杂度和通信负载等方面均明显高于 CDA 算法。

目前，针对多源异构数据数据安全融合的研究成果较少，Lin 等人基于分区域层簇式融合的思想，提出了一种适用于多应用场景的数据融合算法 CDAMA[9]。该方案给部署在区域中的不同类别的传感器节点分配不同的公钥，融合节点可以对不同应用场景下传感器节点采集到的数据进行融合，基站接收到融合数据后，可以根据具体需求使用不同的密钥提取不同应用场景下的数据。但该方案采用了基于椭圆曲线的公钥加密，具有较高的算法复杂度；且它具有较大的密文扩张，显著增加了网络的通信开销。这些问题使其在无线传感器网络资源受限的环境下的实用性大大降低。

针对以上问题，本文提出了一种多应用场景下可同时保障多源异构数据机密性和完整性的安全数据融合方案 SDAMA（Secure Data Aggregation scheme for Multiple Applications）。该算法在明文被加密前对其进行编码，使方案满足多应用场景下的使用需求；以融合轮数 t 作为密钥种子，在实现同态加密的同时保证了每个融合周期中数据的新鲜性；同时，利用同态消息认证码，使得基站能对融合结果进行完整性校验。与已有的多应用安全融合方案 CDAMA 相比，本方案产生的密文扩张更小，通信开销更低，融合精确度更高。

3　背景知识与网络模型

3.1　同态加密

定义 E 为加密函数，m_1，m_2 为明文，$\oplus$ 为明文域上的某种运算，$\otimes$ 为密文域上的某种运算，如果存在一种加密函数 E，使得：

$$E(m_1 \oplus m_2)=E(m_1)\otimes E(m_2)$$

则称 E 是具有“$\oplus$”同态的同态加密算法。当 $\oplus$ 代表加法时，称该加密为加法同态加密；当 $\oplus$ 代表乘法时，称该加密为乘法同态加密。同态加密允许在不解密的情况对密文进行操作，解密后仍然能得到与明文做相应运算后相同的结果。

3.2　同态消息认证码

为了验证数据的完整性，一些单向散列函数（如 MD5，SHA1）通常被用来生成消息认证码。然而，这些消息认证码不能进行有效的融合，会给网络增加额外的通信开销。本文引入了一种轻量级同态消息认证码，计算方法如下：

$$tag_i=f(c_i,k,k_i,P)=(c_i\cdot k)+k_i(\mathrm{mod}\ P) \tag{1}$$

其中，P 为大素数，c_i 为节点感知数据加密后的密文，k 为全局密钥，k_i 为节点与基站共享的密钥。由式（1）为每个密文生成一个认证标签 tag_i。显然，该公式满足加法同态：

$$\begin{aligned} tag_1+tag_2 &= f(c_1,k,k_1,P)+f(c_2,k,k_2,P)\\ &= k\cdot(c_1+c_2)+(k_1+k_2)(\mathrm{mod}\ P)\\ &= f(c_1+c_2,k,k_1+k_2,P)\end{aligned}$$

因此，该消息认证码具有加法同态性。融合节点收到每个成员节点的认证标签 tag_i 后，可以对多个认证标签进行融合，然后上传给基站用于验证融合数据的完整性。

3.3　网络模型

基于簇的数据融合协议的优点是便于管理，降低了节点协作的复杂性，可扩展性好，适合大规模网络。基于簇的数据融合方式将整个网络组织成若干个簇区域，每个区域选举出自己的簇头，传感器节点监测到数据后将数据直接发送到它所在的簇的簇头节点，簇头节点对簇内数据进行融合处理后，转发给 Sink 节点。根据簇头的父亲节点的类型不同，通常有两种簇结构[10]：如果簇头节点和 Sink 节点之间只有一跳距离，则这个簇结构为一跳簇结构；另一种结构为多跳簇结构，每个簇头的父亲节点为另一个簇的成员，数据通过多跳的形式传输至 Sink 节点。

本方案采用第一种分簇方式，即一跳簇结构，如图 1 所示。在这种分簇结构中，网络节点可以分为 3 种基本类型：基站(BS)、融合节点(簇头节点 CH)以及簇内成员节点。由于本文针对多应用场景进行研究，因此设定网络中每个成员节点将会集成多种类型传感器，其负责采集不同种类型的数据(如湿度、温度、压力等)。簇内成员节点负责采集数据并将感知的数据上传给融合节点，如节点 A 将会采集到不同应用场景下的数据，并将这些数据封装在同一个数据包内上传至 Cluster1 的簇头节点；融合节点将簇内所有成员节点发送的数据进行融合后上传给基站节点；BS 网络负责接收簇头节点的融合结果。本文假设基站节点拥有足够的计算和存储能力，且能量不受限制。

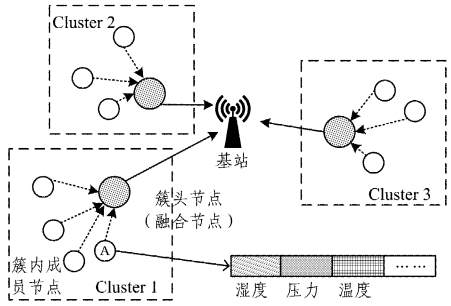

图 1 基于簇的多应用场景的网络拓扑示意图

另外，在安全性方面，本文主要考虑两种攻击类型：1)攻击者通过捕获 WSNs 中的某些传感器节点，从而获得节点密钥等隐私数据，实施窃听攻击；2)攻击者对传感器节点上传的隐私信息进行篡改，实施完整性攻击。

4 算法描述

本方案主要由 4 个阶段构成，分别为初始化阶段、数据加密与标签生成阶段、数据融合阶段和完整性验证阶段。在数据加密与标签生成阶段中，簇内成员节点对采集的数据加密并生成相应的认证标签发送给簇头节点。在数据融合阶段，簇头节点负责将收集到的密文和标签进行加法融合后上传给基站。最后，基站收到各簇头发送的数据包后，对融合结果的完整性进行验证。本文中所有参数及其定义如表 1 所列。

表 1 算法参数及其含义

参数	表示含义
SN_{ij}	第 j 个簇内的传感器节点 i
t	融合周期轮次
‖	连接操作
$m_{ij,\alpha,t}$	SN_{ij} 在第 t 轮周期内采集的数据，数据类型为 α
$d_{ij,t}$	$m_{ij,\alpha,t}$ 经编码后的值
$d_{agg,j}$	$d_{ij,t}$ 的融合结果
$d[a,k]$	提取 $d_{ij,t}$ 的第 a 到 k 个 bit
$c_{ij,t}$	SN_{ij} 在 t 轮周期内感知数据的密文
$c_{agg,t}$	融合密文
k	所有节点与 BS 共享的全局密钥
k_{ij1}，k_{ij2}	每个节点与 BS 共享的密钥
$resID$	所有响应节点 ID 的集合
$PRF()$	使用 SHA-1 作为伪随机函数
$tag_{ij,t}$	第 t 轮周期数据完整性的标签
$tag_{agg,t}$	融合标签

4.1 初始化阶段

在网络被部署之前为每个节点 SN_{ij} 加载密钥 $K_i=(k, k_{ij1}, k_{ij2})$，其中 k 为整个网络中所有节点的共享密钥，k_{ij1} 和 k_{ij2} 为 SN_{ij} 与基站的共享密钥。同时，每个节点需加载一个大数 M、一个大素数 P 和一个伪随机函数 $PRF()$。

4.2 数据加密与标签生成阶段

在本阶段，对于传感器节点 SN_{ij} 采集的数据，本文使用 Castelluccia 等人提出的方案进行加密，并在此基础上引入融合周期计数值 t，基于初始密钥 k_{i1} 和周期数 t 生成密钥 k_{ij1} 和 k_{ij2}。k_{ij1} 和 k_{ij2} 是基于 t 生成的临时密钥，保证了数据的新鲜性，可有效抵御重放攻击。同时，本文在文献[6]中的编码方式的基础上进行了改进，从而满足了多传感数据类型融合的需求。

本文对文献[6]中的编码方式的改进主要分为两个方面：1)在原编码方式中，系统初始化时根据具体数据类型计算出能够表示出该类型数据的最大值所需的最小比特长度 ℓ，假设网络中共有 n 个节点，每个节点除自身感知的 ℓ 比特信息外，还需为其余所有节点都预留 ℓ 比特空位(用 ℓ 个 0 来表示)，最终编码后的数据总长度为 $n\cdot\ell$ 比特，本节点的数据根据节点自身的 ID=i 存放在第 $(i-1)\cdot\ell$ 到第 $i\cdot\ell-1$ 个比特。本文为了实现多源异构数据的融合，对每个节点产生的消息序列不再以 ID 为单位进行区分，取而代之的是每个节点在采集到多种类型数据后，根据数据类型编号将异构数据定位到数据段中的相应位置。2)为了防止同构数据融合产生溢出，本文提出的方案对数据段的格式进行了调整，调整后的格式如图 2 所示。

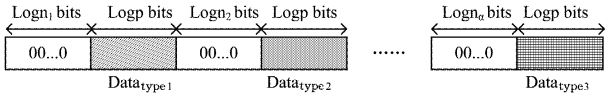

图 2 数据段格式

数据加密与标签生成算法如算法 1 所示。

算法 1 数据加密与标签生成

输入：$m_{ij,\alpha,t}$，k_{i1}，k_{i2}，M

输出：$c_{ij,t}$，$tag_{ij,t}$

1. 编码 $d_{ij}=m_{ij,1,t}\|m_{ij,2,t}\|\cdots\|m_{ij,\alpha,t}$
2. 计算当前数据加密密钥 $k_{ij1,t}=PRF(k_{ij,1},t)$
3. 计算当前认证标签生成密钥 $k_{ij2,t}=PRF(k_{i2},t)$
4. 计算 $c_{ij,t}=d_{ij,\alpha,t}+k_{ij1,t}(\mathrm{mod}\ M)$
5. 计算 $tag_{ij,t}=(c_{ij}\cdot k)+k_{ij2,t}(\mathrm{mod}\ P)$

其中，$|m_{ij,\alpha,t}|=\lambda(\alpha-1)$，$\lambda$ 为能足够表示每一种类型的数据所需的比特数，$\lambda=\log(n_\alpha\cdot p)$，$n_\alpha$ 为网络中第 $\alpha-1$ 种传感器的数量，p 为第 $\alpha-1$ 种传感器感知的信息的最大长度。

4.3 数据融合阶段

在本阶段，簇头节点作为融合节点，使用加法同态操作将簇内成员节点上传的密文和认证标签融合成单个密文和标签。CH 将融合的密文 $c_{agg,j}$ 和标签 $tag_{agg,j}$ 上传给基站。融合过程如算法 2 所示。

算法 2　数据融合

输入：$c_{ij,t}$，$tag_{ij,t}$

输出：$c_{agg,j}$，$tag_{agg,j}$

1. 对于 n 个密文（$c_{1j,t}$，…，$c_{nj,t}$）

计算 $c_{agg,j}=\sum c_{ij,t} \bmod M$

2. 对于 n 个认证标签（$tag_{1j,t}$，…，$tag_{nj,t}$）

计算 $tag_{agg,j}=\sum tag_{ij,t} \bmod P$

4.4　完整性验证阶段

基站在接收到各簇头节点发送的数据包后，首先使用与节点共享的密钥 k_{ij1} 和 k_{ij2}，根据当前融合轮数 t 生成当前网络中所有节点使用的密钥 $k_{ij1,t}$ 和 $k_{ij2,t}$。其次，基站使用密钥 k 和 $k_{ij2,t}$ 根据融合结果重新计算 $tag'_{agg,j}$，并与接收到的融合标签进行比对。若 $tag_{agg,j}=tag'_{agg,j}$，数据包通过完整性验证，则基站继续对密文数据包解密，再根据相应的解码规则提取出不同类型数据的融合值；反之，若 $tag_{agg}\neq tag'_{agg}$，则基站丢弃该融合数据包。验证过程如算法 3 所示。

算法 3　端到端完整性验证

输入：$c_{agg,j}$，$tag_{agg,j}$

输出：验证结果

1. 计算 $k_{ij1,t}=PRF(k_{ij1},t)$

2. 计算 $k_{ij2,t}=PRF(k_{ij2},t)$

3. 计算 $K_{2,t}=\sum_{i\in resID} k_{ij2,t}$

4. 计算 $tag'_{agg,j}=(c_{agg,t}\cdot k)+k_{i2,t} \pmod P$

如果 $tag'_{agg,j}=tag_{agg,j}$

计算 $d_{agg,j}=c_{agg,j}-\sum_{i\in resID} k_{ij,t}$

译码 $m_{agg,\alpha,j}=d[(\alpha-1)\cdot\lambda,\alpha\cdot\lambda-1]$

否则

基站丢弃该数据包

下面通过简单的例子来说明 SDAMA 的工作流程。假设网络中有 3 个节点$\{SN_{11},SN_{21},SN_1\}$，其中 SN_1 为簇头节点。假设每个节点集成 3 种类型的传感器，采集到的数据分别为：

$$\begin{cases} m_{11,1}=3, m_{11,2}=5, m_{11,3}=7 \\ m_{21,1}=4, m_{21,2}=6, m_{21,3}=8 \end{cases}$$

首先，节点对明文编码，此时 $\lambda=4$，编码后的结果为：

$$\begin{cases} d_{11}=(011101010011)_2 \\ d_{21}=(100001100100)_2 \end{cases}$$

然后，节点对编码结果进行加密得到密文 c_{11}，c_{21}，并使用密文生成认证标签 tag_{11}，tag_{21}。随后，SN_{11} 和 SN_{21} 将（c_{11}，tag_{11}）和（c_{21}，tag_{21}）发送给 SN_1。SN_1 收到密文和标签后，分别对其执行加法融合操作，可得：

$$\begin{cases} c_{agg,1}=c_{11}+c_{21} \pmod M \\ tag_{agg,1}=tag_{11}+tag_{21} \pmod P \end{cases}$$

随后 SN_1 将融合结果（$c_{agg,1}$，$tag_{agg,1}$）发送给基站。最后，基站在收到（$c_{agg,1}$，$tag_{agg,1}$）后，先验证数据的完整性，如果验证通过，那么基站对融合密文解密得到 $d_{agg,1}=(111110110111)_2$，再根据译码规则计算：

$$m_{agg,1,1}=d[(1-1)\cdot 4,1\cdot 4-1]=d[0,3]=(0111)_2=7$$

$$m_{agg,2,1}=d[(2-1)\cdot 4,2\cdot 4-1]=d[4,7]=(1011)_2=11$$

$$m_{agg,3,1}=d[(3-1)\cdot 4,3\cdot 4-1]=d[8,11]=(1111)_2=15$$

该结果即为 3 种不同类型数据的最终融合结果。

5　算法分析与仿真

本节从安全性、通信量、融合精确度和多应用场景下的适用性 4 个方面分析 SDAMA 方案的性能，并将其与多应用场景下的数据融合方案 CDAMA 以及基于簇的数据融合方案 iCPDA 方案[11]进行了比较。使用 TinyOS 集成的 TOSSIM 进行仿真实验，仿真参数如表 2 所列。

表 2　仿真参数

仿真参数	取值
背景噪声	−105dBm
高斯白噪声	4dB
节点数量	60～100
分布区域	50m＊50m

5.1　安全性分析

首先，从数据的机密性、完整性和新鲜性 3 个方面对方案的安全性进行分析。

5.1.1　数据机密性

本文考虑如下两种情形：1）攻击者未俘获任何节点，仅对无线信道实施窃听；2）攻击者俘获了一个或多个节点。在第一种情形下，攻击者因缺少密钥 $k_{ij1,t}$ 不能对密文实施解密操作。在方案中，节点每个融合周期使用的密钥由初始密钥 k_{ij1} 和融合轮数 t 通过伪随机函数 $PRF()$ 生成，密钥大小为 20byte。因此，密钥被破解的概率为 2^{-160}，此概率可忽略不计。在第二种情形下，攻击者可以通过俘获节点来获取存储在节点中的密钥信息，进而向网络注入伪造数据包。由于每个节点和基站共享不同的密钥，因此单个或多个节点被俘获并不会危及网络中的其他节点。如果被俘获的节点为融合节点，因为在这种情况下融合节点不存储任何密钥，所以攻击者等同于对信道实施窃听。综上所述，本方案在节点未被俘获的情况下可保证数据的机密性。

5.1.2　数据完整性

通过对数据进行同态加密，保证了数据端到端的机密性，但基站无法判断出数据是否在传输过程中被篡改。本文使用同态消息认证码为每个密文信息生成一个认证标签 $tag_{ij,t}$。如果数据在传输过程中被篡改，基站将检测到根据融合数据包重新计算出的认证标签 $tag'_{agg,j}$ 与原标签值不匹配。另外，该标签满足加法同态，多个认证标签可在融合节点通过加性融合为单一认证标签 $tag_{agg,j}$，从而减少网络中的通信开销。

5.1.3　数据新鲜性

本文中数据周期性地进行融合，基站为了保证收到的融合结果全部是当前周期内产生的，必须对数据的新鲜性进行验证。在本方案中，每个节点加密密钥 $k_{ij1,t}$ 由初始密钥 k_{ij1} 和融合轮数 t 通过伪随机函数 $PRF()$ 生成，密钥随融合轮数 t 动态更新。如果攻击者重放上一轮周期中的数据，基站在收到融合数据后会因使用的加解密密钥不一致导致解密错误，

进而无法通过完整性验证。

5.2 通信量

在 CDAMA 方案中，感知数据明文长度为 $|m|=8$bit，其密文大小和网络中的数据类型数量 k 有关，这里取 $k=3$，则密文大小约为 96byte，Tinyos 协议栈限制 MAC 层数据负载最大为 30byte，每个密文数据需要分为 4 个数据包进行发送。假设网络中共有 n 个节点，则在 CDAMA 方案中网络中总数据包发送个数为 $4n$。

在 iCPDA 方案中，簇内成员节点需要将采集到的数据进行分片并加密发送给其他成员，分片数量与簇的大小有关。由文献[11]可知，网络中每个节点成为簇头的概率为 p，因此平均每个成员节点需要发送的数据包个数为 $1/p$，为保证数据的完整性，该方案需要额外发送 $2pn$ 个汇报数据包。综上，在 iCPDA 方案中，网络中发送的数据包总数约为 $n/p-n+3pn$。

本方案中每个节点需要发送两个数据包，每个数据包的大小为 20byte，在一个融合周期内，整个网络中发送的数据包个数为 $2n$。图 3 给出了 3 种不同方案下网络中总数据包量和节点数量之间的关系。由图 3 可以看出，与 iCPDA 和 CDAMA 方案相比，本方案的数据包发送个数分别减少了近 56% 和 50%。

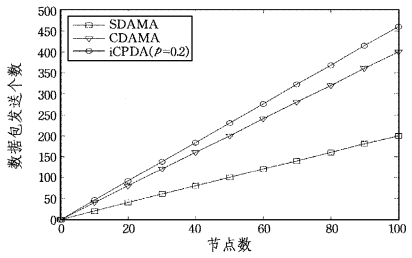

图 3 数据包发送量比较

图 4 给出了当节点数量分别为 60，80，100 时一个融合周期中 3 种方案分别在网络中产生的总数据通信量。仿真结果表明，本文提出的方案具有较小的通信开销。由于 SDAMA 使用的是对称加密，与 CDAMA 相比其产生的密文较短，使得整个数据包的长度也较短。与 iCPDA 相比，SDAMA 无需对数据进行分片，因此每个节点需要发送的数据包数量相对较少。

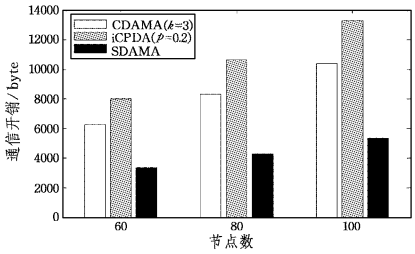

图 4 不同节点密度下的通信开销比较

5.3 计算开销

本节对 SDAMA，iCPDA 和 CDAMA 3 种方案的簇内成员节点的计算开销进行分析，所使用的计算开销符号如表 3 所列。

表 3 计算开销符号及其定义

符号	定义
C_{PA}	椭圆曲线上点加计算开销
C_M/C_A	模乘/模加计算开销
C_{PRF}	伪随机函数的计算开销
C_E/C_D	加密/解密数据的计算开销

根据文献[9]，在 CDAMA 方案中，当应用数量 $k=3$，数据明文长度为 $|m|=8$bit，随机数长度 $|r|=80$bit 时，每个节点加密产生的计算开销约等于 $1+\frac{3|m|}{2}+\frac{3|r|}{2}=133$ 次椭圆曲线上的点加运算，因此簇成员节点的计算开销可表示为：

$$C_{CDAMA}\approx 133C_{PA}$$

在 iCPDA 方案中，簇内成员节点的分片数量取决于节点成为簇头的概率 p，取 $p=0.2$，则每个簇平均有 $\frac{1-p}{p}=4$ 个成员，分片数量为 4，每个节点需要发送和接收 3 个数据分片并分别对其进行加解密操作。因此，簇成员节点的计算开销为：

$$C_{iCPDA}=3C_A+3(C_E+C_D)$$

在本方案中，各簇内成员节点需要执行 2 次 $PRF(\cdot)$，用于生成一个融合周期的数据加密密钥和认证标签生成密钥；一次模加操作用于数据加密；一次模加和模乘用于生成认证标签。因此，SDAMA 方案的簇成员节点的计算开销为：

$$C_{SDAMA}=2C_{PRF}+C_M+3C_A$$

由于模加、模乘操作的计算开销远小于 CDAMA 中椭圆曲线上的点加运算和 iCPDA 中的加解密复合运算，因此综合上述理论分析可知，相比于其他两种方案，SDAMA 方案的簇成员节点的计算开销更小。

5.4 精确度

精确度是衡量数据融合算法性能的一项重要指标。数据在实际传输过程中受到信道噪声、数据发送时延以及数据碰撞等因素的影响，数据融合精确度无法达到 100%。图 5 给出了 SDAMA，CDAMA 和 iCPDA 3 种方案在不同融合周期下的精确度。仿真结果显示，融合精确度随着融合周期的变长而提高。iCPDA 在数据融合之前首先要对原始数据进行分片。因此，在融合周期较短时，由于留给数据分片和上传的时间不足，导致其精确度较低。随着融合周期变长，时间逐渐成为影响精确度的次要因素。在 CDAMA 和 iCPDA 方案中，每个节点要发送多个数据包，因此会增大数据包发生碰撞和丢失的概率，进而降低数据融合的精确度。图 6 给出了本方案在 60，80，100 个节点下精确度与融合周期的关系。结果表明，融合精确度会随节点密度的增加而下降。

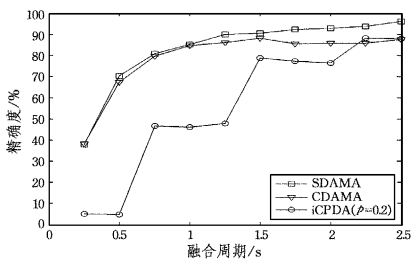

图 5 SDAMA，CDAMA 和 iCPDA 的精确度比较

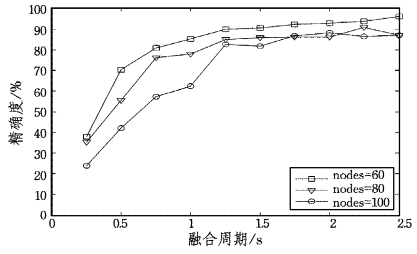

图 6　SDAMA 在不同拓扑下的精确度

5.5　多应用场景下的适用性

在 3 种方案中，iCPDA 只支持单一的数据类型，不适用于多应用场景。CDAMA 理论上支持的应用数量没有上限，其密文长度与应用数量 k 有关。由文献[8]可知，当应用数量 $k=4$ 时，8bit 的明文加密后产生的密文长度已经达到 1280bit。继续对其进行扩展可满足更多应用场景下的使用需求，但同时密文扩展也使得网络中的通信开销大幅增加。在 SDAMA 方案中，应用支持数目与网络规模以及单个节点感知的明文信息的最大长度有关。假设单个节点采集到每种数据类型的最大长度为 8bit，在密文长度不变的情况下，网络中支持的应用数量 k 与节点数量 n 的关系如图 7 所示。当节点数达到 600 时，网络中允许的最大应用数目仍可达到 13 个。

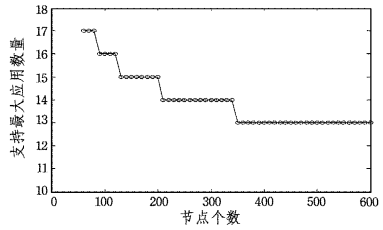

图 7　SDAMA 节点数与最大应用数量的关系

表 4 对比了 3 种方案在特定参数设定下网络中允许的最大应用数量。对比结果表明，SDAMA 在数据包长度较小的情况下可获得对更多应用数量的支持。

表 4　3 种方案在多应用场景下的适用性对比

方案	数据段负载	支持应用数量
SDAMA(n=600)	20byte	13
CDAMA(k=3)	96byte	3
CDAMA(k=4)	160byte	4
iCPDA	不支持多应用场景	

结束语　针对多应用场景下多源异构数据的安全融合问题，本文采用加法同态加密和同态消息认证码分别为数据提供了端到端的机密性和完整性保护。理论分析和仿真结果表明，与 CDAMA 和 iCPDA 相比，SDAMA 具有较小的通信开销和较高的融合精确度。同时，在多应用场景下，与 CDAMA 相比，SDAMA 在不增加额外通信开销的情况下支持的应用数量更多，比较适用于能量资源有限的传感器网络。由于本方案基于加法融合，基站无法从最终融合结果中提取出每个节点的原始数据，因此在融合函数的使用上比较受限。在未来的工作中将研究数据可恢复机制以及恶意节点检测机制，使网络能够尽早检测并丢弃虚假数据包，以减少节点的能源消耗。

参 考 文 献

[1] XU J, YANG G, CENG Z Y, et al. A survey on the privacy-preserving data aggregation in wireless sensor networks [J]. China Communications, 2015, 12(5): 162-180.

[2] SIRSIKAR S, ANAVATTI S. Issues of Data Aggregation Methods in Wireless Sensor Network: A Survey [J]. Procedia Computer Science, 2015, 49: 194-201.

[3] GAIKWAD P B, DHAGE M R. Survey on Secure Data Aggregation in Wireless Sensor Networks [C] // Proceedings of the 2015 International Conference on Computing Communication Control and Automation (ICCUBEA). Piscataway, NJ: IEEE, 2015: 242-246.

[4] ZHOU Q, YANG G, HE L W. An Efficient Secure Data Aggregation Based on Homomorphic Primitives in Wireless Sensor Networks [J]. International Journal of Distributed Sensor Networks, 2014, 2014(7): 38-50.

[5] CLAUDE C, MYKLETUN E. Efficient and provably secure aggregation of encrypted data in wireless sensor networks [J]. ACM Transactions on Sensor Networks, 2009, 5(3): 1137-1153.

[6] 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.

[7] CHEN C M, LIN Y S, LIN Y C. RCDA: recoverable concealed data aggregation for data integrity in wireless sensor networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(4): 727-734.

[8] VERMA S, PILLAI P, HI Y F. Energy-efficient privacy homomorphic encryption scheme for multi-sensor data in WSNs [C] // 2015 7th International Conference on Communication Systems and Networks (COMSNETS). Piscataway, NJ: IEEE, 2015: 1-6.

[9] LIU Y S, CHANG S Y, SUN H M. CDAMA: Concealed Data Aggregation Scheme for Multiple Applications in Wireless Sensor Networks [J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(7): 1471-1483.

[10] CHEN Z Y. Research on QoS-Oriented Data Collection for Wireless Sensor Networks [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2015. (in Chinese)
陈正宇. 面向 QoS 的无线传感器网络数据收集方法研究 [D]. 南京：南京邮电大学，2015.

[11] HE W B, LIU X, NGUYEN H V. PDA: Privacy-Preserving Data Aggregation for Information Collection [J]. Acm Transactions on Sensor Networks, 2011, 8(1): 108.