计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2016.20150456
53(9):2030-2038,2016

一种无线传感器网络隐私保护数据聚合方案

付 帅[1,2] 姜 奇[2] 马建峰[2]
[1](北京电子科技职业学院电信工程学院 北京 100176)
[2](西安电子科技大学计算机学院 西安 710071)
(ljhfs0803@126.com)

A Privacy-Preserving Data Aggregation Scheme in Wireless Sensor Networks

Fu Shuai[1,2], Jiang Qi[2], and Ma Jianfeng[2]
[1](*School of Electronic and Information Engineering*, *Beijing Polytechnic*, *Beijing* 100176)
[2](*School of Computer Science and Technology*, *Xidian University*, *Xi'an* 710071)

Abstract Privacy preservation is one of the most challenging problems on secure data aggregation in wireless sensor networks (WSNs). In WSNs, current data aggregation schemes with privacy preservation cannot meet the requirements of security and energy saving, and have several disadvantages such as complex computation, considerable communication load or low-security. An energy-efficient and data-loss resilient data aggregation scheme with privacy preservation is proposed in this paper. Two different forms of perturbation data are adopted to protect the data privacy of each node from being disclosed to the sink and any other nodes in the network. Firstly, from the point of view of sink intrusion, we describe the design scheme of initial perturbation data. On the basis of it, we present the construction method of second data perturbation and the operation procedures of aggregation validation for intermediate aggregators and the sink. To resist various external attacks efficiently, the technique of message authentication code is introduced. Security and property analysis show that the proposed scheme can ensure the security of nodes on the premise of lower energy power. In addition, it has a strong ability against data-loss, and both its security and energy efficiency perform better than current works.

Key words wireless sensor networks (WSNs); data aggregation; privacy preserving; data perturbation; energy efficient

摘 要 隐私保护是基于无线传感器网络(wireless sensor networks, WSNs)的数据聚合技术中最具挑战性的安全问题之一.在WSNs环境中,现有的隐私保护数据聚合机制不能同时满足安全性及节能性要求,存在计算复杂、通信量大及安全性低等缺点.提出一种能量有效的、抗数据丢失的隐私保护数据聚合方案,该方案利用2次不同形式的数据扰动同时实现了数据对基站及网内其他节点的隐私保护.

收稿日期:2015-06-09;**修回日期**:2015-11-13
基金项目:长江学者和创新团队发展计划基金项目(IRT1078);国家自然基金委员会-广东联合基金重点基金项目(U1135002);国家自然科学基金项目(61272541,61202389);中央高校基本科研业务费专项资金(JY10000903001);陕西省自然科学基础研究计划基金项目(2012JQ8043);北京市教委教改课题(PXM2015_014306_000017)
This work was supported by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1078), the Key Program of NSFC-Guangdong Union Foundation (U1135002), the National Natural Science Foundation of China (61272541, 61202389), the Fundamental Research Funds for the Central Universities (JY10000903001), the Natural Science Basic Research Plan in Shaanxi Province of China (2012JQ8043), and the Teaching-Reform Project of Beijing Municipal Education Commission (PXM2015_014306_000017).

首先,从防止基站入侵角度,给出了初次扰动数据设计方法;在此基础上,为实现对邻居节点的隐私保护,提出二次扰动数据的构造方法,并给出中间聚合节点及基站的聚合验证操作流程.通过引入消息认证码技术,有效抵御了多种外部攻击.安全及性能分析表明,该方案可在不过多消耗节点能量的前提下保证节点的安全性,且具有较好的抗数据丢失能力,安全性及能效性均优于现有方案.

关键词　无线传感器网络;数据聚合;隐私保护;数据扰动;能量有效

中图法分类号　TP393

在无线传感器网络(wireless sensor networks, WSNs)中,由于传感器节点能量有限,以减小传输数据量、延长网络生命周期为目的的数据聚合技术常应用在 WSNs 中[1].而在 WSNs 早期的研究中,传感器节点通常被部署在边境、战场等无人值守、复杂恶劣的环境中,因而大多不考虑人为因素,WSNs 中数据的隐私安全问题也一直未得到充分的关注.但是,随着研究的深入和技术的不断成熟,WSNs 逐渐扩展到交通管理、智能家居等诸多与人类日常生活密切相关的领域[2].传感器节点能够采集和记录各种活动数据,且其采集的数据往往携带重要的敏感信息.在数据聚合过程中,中间节点需要对接收的数据进行处理并向上传递,从而可能获取到相应的敏感信息.因此,人类在享受高科技带来的便利及周到服务的同时也面临着一场前所未有的隐私危机.目前,WSNs 隐私问题已引起业界的广泛关注,保护人类的隐私安全、保证被监测目标的敏感信息不被未授权者获取具有重要的研究意义.但是,隐私保护技术的应用在一定程度上增大了数据聚合机制的复杂性,因而造成节点能耗量的增加.所以,能量有效性成为一项在数据聚合过程中与隐私保护紧密相关又相互制约的问题.

为实现数据聚合机制的隐私性、能效性及健壮性,一个高效的隐私保护聚合方案应同时满足以下安全及性能要求:1)隐私保护的有效性.首先,能够确保每个节点的原始数据对基站或网络中其他任意节点的隐私性;其次,在网络中部分节点或基站被捕获、且攻击者能够窃听所有通信信息的情况下,高效的隐私保护数据聚合方案仍可以最大程度地降低未被捕获节点的敏感数据泄露概率.2)算法效率.在资源受限的 WSNs 中,数据聚合机制的应用旨在降低网络能耗及节省通信带宽.高效的隐私保护数据聚合算法应在无明显增加通信负载及计算代价的前提下达到隐私保护的目的.3)抗数据丢失能力.在 WSNs 中,数据丢失意味着部分节点未参与聚合.因此,对于一个高效的隐私保护数据聚合方案,基站应可以在移除与隐私保护相关的信息后正确提取出所有参与聚合的节点的最终聚合结果,不会因部分数据丢失而导致无法正确计算最终聚合值.目前,已有的典型数据聚合隐私保护机制并不能同时满足以上安全及性能要求[3-6].文献[3]提出了一种针对加性聚合函数 *sum* 的隐私保护数据聚合方案 PDA(privacy-preserving data aggregation),具体包括 CPDA (cluster-based private data aggregation)和 SMART (slice mix aggregate)两种算法.CPDA 是基于分簇的数据聚合算法,该算法计算过程复杂且计算量大.SMART 则采用切分重组技术实现对聚合数据的隐私性保护:通过将私有数据切分为 J 个数据切片(piece)并分发给随机选择的邻居节点来隐藏原始数据,其通信开销和保护力度均随 J 的增长而增长.文献[4]采用了一种简单的求和同态加密函数(additively homomorphic encryption, AHE).该机制存在以下缺陷:假如攻击者获取了隐藏数据及秘密数的范围,则能够推测出相应原始数据的范围.另外,该机制不能满足聚合数据对基站的隐私性,这一问题在基站妥协时尤为突出.文献[5]提出的 ESPART (energy-saving privacy-preserving data aggregation)方案借助了树形结构本身的特性达到节省能耗、降低通信量的目的,延长了网络的寿命.文献[6]针对 PDA 方案计算量及通信量大的特点进行了改进,提出了一种基于复数代数表达式的隐私保护方案.这 2 种典型的隐私保护技术能够保证聚合数据的安全且提高了算法的效率,但并未对其抗数据丢失能力进行评估.针对上述问题,本文提出了一种能量有效的隐私保护数据聚合方案,可在保证算法效率的前提下保护节点的数据隐私,且健壮性高.

1　预备知识

1.1　消息认证码

本文采用消息认证码机制为数据完整性的实现提供保证.定义 $h(\cdot)$ 表示一个安全的加密 Hash

函数,则依据文献[7]可在$\mathbb{Z}_{2^n}$中构造带密钥的 Hash 函数 $MAC(m,k,n)=h(m\|k) \bmod 2^n$. 其中,$m$,$k$,$n$ 分别表示消息内容、密钥和可调整参数. 若 $n=1$,则 $MAC(m,k,1)$可提供 1 b 的认证,即过滤错误消息的概率为 1/2;若 $n=\theta$,则 $MAC(m,k,\theta)$可以更高的概率 $1-\frac{1}{2^\theta}$过滤错误数据.

1.2 密钥分配机制

为每个节点分配 2 种类型的密钥:

1) 基于 TinyECC 的非交互式密钥对. 假设任一大素数 p,$E(F_p)$为定义在有限域 F_p 上的椭圆曲线,设 G 为 $E(F_p)$中阶为素数 q 的基点. 为每个传感器节点 N_i 加载一个基于 TinyECC 的公私钥对 (Y_i,x_i),其中,私钥 $x_i\in\mathbb{Z}_q^*$,公钥 $Y_i=x_iG$. 对于任意 2 个传感器节点 N_i,$N_j\in G(V,E)$,无论 N_i 和 N_j 能否直接通信,持有密钥对(Y_i,x_i)的节点 N_i 和持有密钥对(Y_j,x_j)的节点 N_j 都能建立一个安全的椭圆曲线 Diffie-Hellman(ECDH)密钥对[8]:

$$k_{ij}=x_iY_j=x_ix_jG=x_jx_iG=x_jY_i=k_{ji}. \quad (1)$$

基于椭圆曲线离散对数问题的困难性,只有节点 N_i 和 N_j 可以共享同一密钥. 同时,建立的密钥对是独立的,即若节点 N_i 妥协,只会泄露 N_i 和 N_j 共享的密钥 k_{ij},而 N_j 和其他节点共享的密钥 k_{jx} 不受影响. 设本文中每个节点 N_i 都与父节点 N_p 建立了相应的安全 ECDH 密钥对,分别应用在聚合过程的不同阶段.

2) 基于随机密钥的分配方法[9]. 产生一个具有 X 个密钥的密钥池,每次基站发起聚合请求时,每个节点都会从该密钥池中随机选取 α 个密钥 k_1,k_2,…,k_α 作为私钥. 因此,一个节点可能与其他节点匿名共享密钥池中的部分密钥. 在 $x\ll\alpha$ 时,任意 2 个节点匿名共享完全相同的私密钥的概率非常小. 在本文所提的聚合算法中,因构造扰动数据需要,应确保聚合过程中所选择的任何密钥都被至少 1 对节点匿名共享.

2 网络及攻击模型

2.1 网络模型

假设 N 个传感器节点随机分布在一个面积为 S 的区域内,且该传感器网络具有如下性质:1)网络是静态的,节点部署后不再移动,且每个节点拥有一个唯一的非零标识符(ID);2)链路是对称的,所有节点都可通过无线信道和邻居节点进行通信;3)基站位置是固定的,且具有强大的计算能力和充足的能量. 可将该传感器网络抽象为连通图 $G(V,E)$,其中,$|V|=|\{N_1,N_2,\cdots\}|=N$ 表示传感器的节点数;$E=\{(N_i,N_j)\,|\,N_i,N_j\in V\}$为节点间的通信链路. 如图 1 所示,本文中的传感器网络采用树形结构. 因重点关注数据聚合过程中的安全问题,且现已有诸多文献对 WSNs 中构建路由树问题进行了大量研究并取得了系列成果,所以不再对其进行详细讨论.

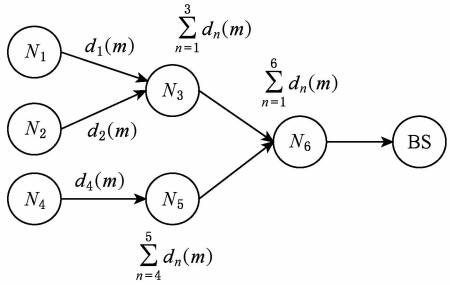


Fig. 1 Data aggregation mode in WSNs.

图 1 传感器网络数据聚合模型

在聚合开始阶段,基站首先向距离最近的根节点发出聚合请求消息,再由根节点通过孩子节点逐级向下发送到所有节点. 聚合响应过程数据的发送顺序则正好相反,由叶子节点开始,将数据通过父节点逐级向上传送到基站.

典型的聚合函数有求最大、最小、均值及方差等,但这些函数均可以转化为求和函数,所以求和是一种基础聚合函数,实现求和函数的隐私保护则意味着可以实现该系列函数中的隐私保护. 本文仅讨论求和函数. 定义加性聚合函数

$$y(t)=f(d_1(m)+d_2(m)+\cdots+d_N(m)),$$

其中,$d_i(m)(i=1,2,\cdots,N)$表示节点 i 在第 m 次聚合过程中采集到的数据.

2.2 攻击模型

1) 隐私性攻击. 外部攻击者试图通过窃听等手段捕获节点间无线传输的信息,从而推测获取其隐私数据所对应的明文信息. 网内邻居节点或基站遵循“诚实但好奇”模型. 另外,如果网络中的部分节点被攻击者捕获,攻击者将会试图利用妥协节点的私密信息推测原始数据;而当大量节点被攻陷时,攻击者则会利用这些妥协节点发起共谋攻击,试图获取网络中的敏感信息.

2) 完整性攻击. 数据聚合过程面临的完整性攻击主要涉及 3 个方面:①外部攻击者发起的注入错误数据攻击;②节点被捕获后,攻击者利用获取的

节点内部信息对消息进行篡改或伪造;③攻击者在捕获节点间无线传输的消息后故意延迟发送发起的重放攻击.我们认为少量被物理捕获的普通传感器节点对网络不构成安全威胁,且仅依赖安全协议很难避免妥协节点的隐私数据不被泄露,所以重点关注外部攻击节点对聚合结果的篡改或伪造,并设法使基站接受该错误消息,不考虑妥协节点对自身感知数据的修改.

3　高效的隐私保护数据聚合方案

为达到 WSNs 中隐私保护有效性、高效性及抗数据丢失能力的要求,本节设计了一种高效的隐私保护数据聚合方案.该方案将 WSNs 数据聚合过程划分为聚合请求、密文构造和聚合及验证 3 个阶段.本节将对每一阶段的实现过程进行详细描述.

3.1　聚合请求

基站每次准备收集节点的感知数据时,首先向距离最近的根节点发送聚合请求消息,然后由根节点通过孩子节点逐级向下发送到网络中所有节点.聚合请求及响应过程均采用了构建索引值的思想.基站首先向生成树的根节点发送聚合请求消息$(R_m, IX^{\lambda}_{\mathrm{R,BS}})$.其中,$m$ 表示当前聚合过程的轮数,$IX^{\lambda}_{\mathrm{R,BS}}$表示基站发送的 λ 位请求索引值(λ 为系统参数).设基站的 ID 为 0 且不存储任何密钥,则 $IX^{\lambda}_{\mathrm{R,BS}}=0$.然后,由根节点逐级向下发送$(R_m, IX^{\lambda}_{\mathrm{R},i})$.$IX^{\lambda}_{\mathrm{R},i}$表示节点 N_i 的请求索引值,隐含了其密钥分布情况.基站向根节点发送聚合请求消息的过程可分为 2 个阶段:1)基站设定自身索引值为 λ 位的全 0 字符串 $IX^{\lambda}_{\mathrm{R,BS}}=(0)^{\lambda}$;2)基站向根节点 N_{r} 发送请求消息$(R_m,(0)^{\lambda})$.

根节点在接收到基站的聚合请求消息后重新确定自身索引值,并向所有孩子节点转发该请求.节点聚合请求索引值的计算方法如下:对密钥池中的任意密钥 k_w,1)若节点 N_j 拥有密钥 k_w,则 N_j 将其所有孩子节点索引值的第 w 位设置为 1,以保证其孩子节点在计算返回的聚合结果时可以应用 k_w;2)若 N_j 没有选取 k_w,但其父节点发送的请求消息中索引值的第 w 位为 1(父节点选取了密钥 k_w),则 N_j 将任意选择一个孩子节点,设定其请求索引值的第 w 位为 1;3)若 N_j 及其父节点均未选取 k_w,则 N_j 发往所有孩子节点的请求消息中索引值的第 w 位均为 0.具体算法表述如下:

算法 1. 中继节点请求索引值分配算法.

① N_j 从父节点 N_t 接收聚合请求消息 $(R_m, IX_{\mathrm{R},t})$;
② if N_j 为叶子节点 then
③　return; /* 开始聚合 */
④ end if
⑤ for 密钥池中任意密钥 $k_w(1\leqslant w\leqslant p)$ do
⑥　if k_w 是节点 N_j 的私钥 then
⑦　　for N_j 的所有孩子节点 N_{children} do
⑧　　$IX^{w}_{\mathrm{R,c}}=1$;
⑨　　end for
⑩　else if $IX^{w}_{\mathrm{R},j}=1$ then
⑪　　从孩子节点中随机选择节点 $N_q\in N_{\mathrm{children}}$;
⑫　　$IX^{w}_{\mathrm{R},q}=IX^{w}_{\mathrm{R},j}=1$;
⑬　　for N_j 的其他孩子节点 $N_p(N_p=N_{\mathrm{children}}-N_q)$ do
⑭　　$IX^{w}_{\mathrm{R},p}=0$;
⑮　　end for
⑯　end if
⑰ end for
⑱ for N_j 的每个孩子节点 N_{children} do
⑲　N_j 向其发送聚合请求消息 $(R_m, IX_{\mathrm{R},j})$;
⑳ end for

同时,以 $\lambda=5$,$\alpha=2$ 为例,图 2 给出了聚合请求转发过程中索引值的计算示例.

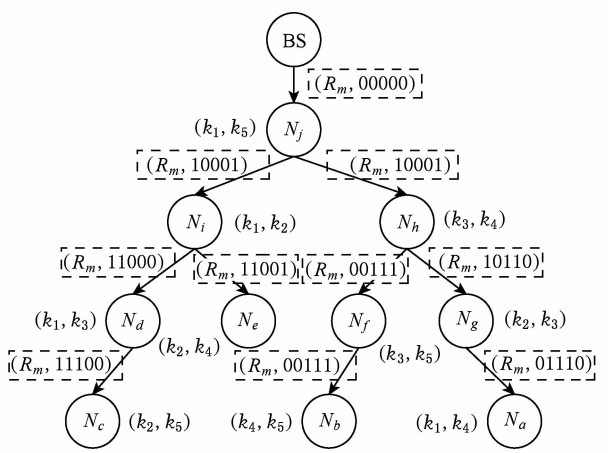


Fig. 2　Construction of the request index value.

图 2　请求索引值计算示例

3.2　密文构造

所有节点均接收到基站发来的聚合请求后,由叶子节点开始,通过父节点逐级向上将感知数据传送到基站.为同时实现节点的原始数据对基站和

网络中其他任意节点的隐私性,每个节点均采用数据扰动的方法对提交数据进行2次不同形式的加密.具体实现过程如下:

1) 基础密文构造.在一些特定环境中,传感器节点采集的数据值往往在某个特定的范围内,如用来采集人体温度的传感器的感知数据值通常在[34℃,42℃].所以,为防止攻击者进行蛮力搜索,首先采用以下数据扰动方法对原始数据进行隐藏:

假定原始数据 d_i 为 l(l 为系统参数,单位:b),节点选择 θ(单位:b)的秘密随机数 r_i,θ 为系统参数,决定了蛮力搜索的难度,得到密文

$$C_i=2^{l+\mathrm{lb}\,N}\times r_i+d_i. \tag{2}$$

执行 *sum* 聚合后,得到的密文数据满足以下特性:

$$\sum_{i\in\mathbf{N}_+}C_i=2^{l+\lfloor\mathrm{lb}\,N\rfloor}\times\sum_{i\in\mathbf{N}_+}r_i+\sum_{i\in\mathbf{N}_+}d_i\leqslant 2^{l+\lfloor\mathrm{lb}\,N\rfloor}\times\sum_{i\in\mathbf{N}_+}r_i+N(2^l-1)<2^{l+\lfloor\mathrm{lb}\,N\rfloor}\times\sum_{i\in\mathbf{N}_+}r_i+2^{l+\lfloor\mathrm{lb}\,N\rfloor},$$

因此,

$$\sum_{i\in\mathbf{N}_+}d_i=\sum_{i\in\mathbf{N}_+}C_i \bmod 2^{l+\lfloor\mathrm{lb}\,N\rfloor}. \tag{3}$$

应用该特性,基站可以在对所有节点选择的随机数未知的情况下计算得到最终的聚合结果.同时,从该特性还可以看出,节点只提交密文 C_l 是不安全的.假定叶子节点 N_l 向其父节点 N_p 发送密文 C_l,N_p 将接收到的密文 C_l 与自身密文 C_p 进行聚合,得到 $C'=C_l+C_p$.利用上述特性,N_p 可获取明文聚合值 d'.然后,通过计算 $d'-d_p$ 即可获取叶子节点 N_l 的原始数据 d_l.所以,只提交密文 C_l 不能满足节点原始数据的隐私性要求.基于此,本文设计了一种构建索引值的数据扰动方案,通过对基础密文添加扰动索引值完成聚合密文的构造,保证了单个节点数据对网络中其他节点和基站的隐私性.

2) 聚合密文构造.每个节点在向其上级节点提交聚合数据前,采用一种构建索引值的方式对基础密文数据进行二次扰动.3.3节将详细描述其实现过程.

3.3 聚合及验证

1) 叶子节点的聚合.在聚合响应阶段,叶子节点 N_l 应用其私钥计算新的 λ 位聚合索引值.每个叶子节点按如下步骤进行操作:①若节点 N_l 拥有密钥 k_w,则 N_l 将其索引值的第 w 位设置为1,其他位设置为0;②N_l 将初步确定的 λ 位索引值与其接收的请求索引值进行与操作,得到聚合索引值 $IX_{\mathrm{A},l}$;③N_l 应用该聚合索引值计算扰动数据 A_w:若索引值的第 w 位为1,则 $A_w=H(R_m,k_w)$.将扰动数据 A_w 与基础密文 C_l 进行运算得到聚合密文 S_l.具体算法表述如下:

算法 2. 叶子节点聚合索引值分配算法.

/* 假定任意叶子节点 N_l */

① for 密钥池中任意密钥 k_w($1\leqslant w\leqslant p$) do

② if k_w 是节点 N_l 的私钥 then

③ $IX^w_{\mathrm{A},l}=1$;

④ else

⑤ $IX^w_{\mathrm{A},l}=0$;

⑥ end if

⑦ end for /* 得到 IX_{A,l_0} */

⑧ $IX_{\mathrm{A},l}=IX_{\mathrm{A},l_0}\wedge IX_{\mathrm{R},l}$;

⑨ for $w=1$ to p do

⑩ if $IX^w_{\mathrm{A},l}=1$ then

⑪ $A_w=H(R_m,k_w)$;

⑫ $S_l=C_l+A_w$;

⑬ end if

⑭ end for

算法1运行结束后,N_l 利用与父节点 N_p 的独立共享密钥 k_{lp} 生成消息认证码 $mac_{lp}=MAC(S_l\parallel IX_{\mathrm{A},l},k_{lp},1)$,并向其父节点发送消息:$N_l\rightarrow N_p$:$S_l\parallel IX_{\mathrm{A},l}\parallel mac_{lp}\parallel T$.其中,$T$ 表示时间戳.

2) 中间节点的聚合.对于中间聚合节点,父节点 N_p 接收到 N_l 的消息后,首先对其完整性进行验证.N_p 使用与 N_l 的独立共享密钥 k_{pl} 计算 mac_{pl}.若 $mac_{pl}\oplus mac_{lp}=0$,则通过验证;否则,认为数据出错而放弃.当 N_p 接收到所有孩子节点的消息并完成验证后,先与自身密文数据 C_p 进行运算得到 S'_p,然后确定其聚合索引值并计算扰动数据.每个中间聚合节点 N_p 遵循如下原则对其索引值进行分配:对密钥池中任意密钥 k_w,①若节点 N_p 拥有密钥 k_w,则 N_p 将其索引值的第 w 位设置为与接收的请求索引值中第 w 位相同的值,并计算扰动数据 A_w;然后,将 A_w 与 S'_p 进行运算得到待发送数据 S_p.②若 N_p 未选取密钥 k_w,则考虑 N_p 的孩子节点个数.如果只有一个孩子节点,则 N_p 聚合索引值的第 w 位取其孩子节点聚合索引值第 w 位的值;否则,N_p 的聚合索引值中第 w 位的取值等于其所有孩子节点第 w 位取值的异或结果.中继节点聚合索引值的具体分配算法表述如下:

算法 3. 中继节点聚合索引值分配算法.

/* 假定中继节点 N_q,N_p 和 N_z 为 N_q 的孩子节点 */

① for N_q 的任意孩子节点 N_p do
② 　N_q 从 N_p 接收消息$\langle S_p, IX_{A,p}\rangle$;
③ end for
④ $S_q=\sum_{N_p} S_p$;
⑤ $S_q=S_q+C_q$;
⑥ for 密钥池中任意密钥 $k_w(1\leqslant w\leqslant p)$ do
⑦ 　if k_w 是节点 N_q 的私钥 then
⑧ 　　$IX^w_{A,q}=IX^w_{R,q}$;
⑨ 　　$S_q=S_q+(IX^w_{A,q}-\sum_{N_p} IX^w_{A,p})A_w$;
⑩ 　else if N_q 只有一个孩子节点 N_p then
⑪ 　　$IX^w_{A,q}=IX^w_{A,p}$;
⑫ 　else
⑬ 　　$IX^w_{A,q}=IX^w_{A,p}\oplus IX^w_{A,z}(N_p,N_z)$;
　　/ * N_p,N_z 为 N_q 的不同的孩子节点 * /
⑭ 　end if
⑮ end for
⑯ if N_q 不是根节点 then
⑰ 　N_q 向父节点发送聚合数据$\langle S_q, IX_{A,q}\rangle$;
⑱ else
⑲ 　根节点向基站发送数据$\langle S_0=S_q\rangle$;
⑳ end if

最后,N_p 生成消息认证码 $mac_{pq}=MAC(S_p \| IX_{A,p}, k_{pq}, 1)$,并向其父节点 N_q 发送消息:$N_p \rightarrow N_q: S_p \| IX_{A,p} \| mac_{pq} \| T$.

结合图 2,图 3 给出了聚合响应过程中索引值及扰动数据的计算示例.

3) 根节点的聚合. 对于基站,当根节点 N_r 将数据 S_0 传送给基站后,基站首先应用 k_{sr} 计算 mac_{sr}. 若 $mac_{sr}\oplus mac_{rs}\neq 0$,认为该数据出错,直接放弃;否则,基站根据接收到的 S_0 值计算其明文聚合结果. 由式(3)可知:

$$\sum_{i\in \mathbf{N}_+} d_i = S_0 \bmod 2^{l+\lfloor \mathrm{lb}N\rfloor}. \tag{4}$$

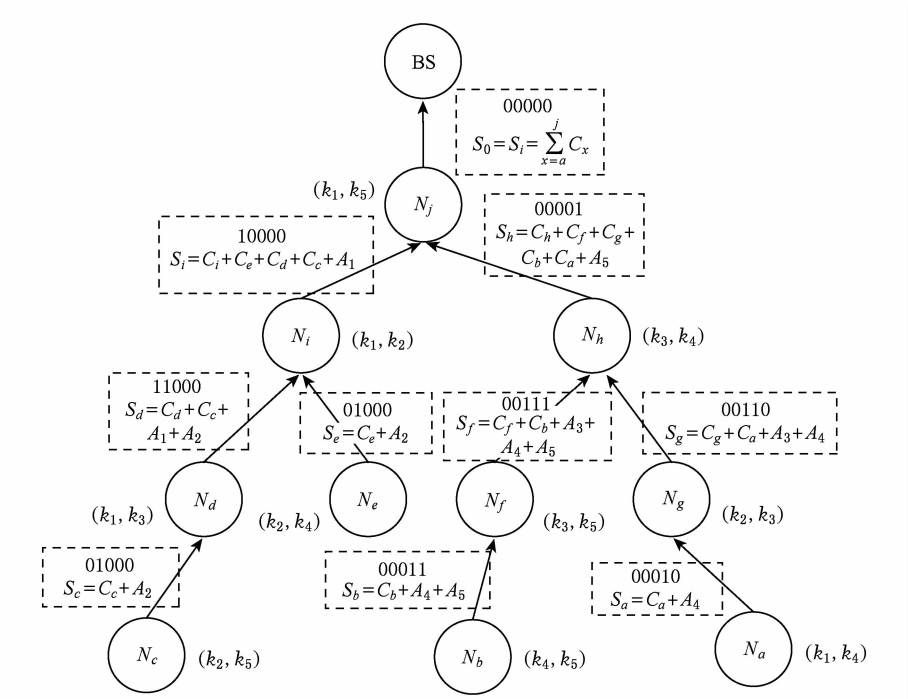


Fig. 3　Construction of aggregation index value and perturbation data.

图 3　聚合索引值及扰动数据计算

4　安全分析

4.1　隐私性

1) 外部节点攻击. 每个传感器节点在向基站发送数据时采用了添加扰动数据的方式进行加密,即 $S_y=C_y+\sum_{y\in\lambda}A_y$ 且 $A_y=H(R_m,k_y)$. 由于任意中间聚合节点不具有节点 y 的私钥 k_y,所以不能推测出基础密文 C_y,保证了单个节点数据对网络中其他节点的隐私性. 基站从根节点接收到消息后,可利用式(3)得到相应的明文聚合结果,而非每个节点的原始数据,保证了单个原始数据对基站的隐私性.

2) 妥协节点攻击. 攻击者可通过捕获传感器节点 N_i 获得该节点的私钥 $k_i(i\in X)$,并试图通过

计算 A_i 推测其他节点 N_j 的原始数据 d_j.假定除目标节点外,攻击者共捕获了 Y 个节点,且对目标节点的所有通信流进行了窃听.用 $Q_i(1\leqslant i\leqslant\alpha)$ 表示攻击者对目标节点的第 i 个私钥是未知的,则目标节点索引值泄露的概率为

$$P_{\text{tar}}=1-P(Q_1\vee Q_2\vee\cdots\vee Q_\alpha)=1-(\sum_{i\in[1,\alpha]}P(Q_i)-\sum_{i_1,i_2\in[1,\alpha],i_1<i_2}P(Q_i\wedge Q_j)+\cdots+(-1)^{\alpha-j-1}\times\sum_{i_1,i_2,\cdots,i_{\alpha-j}\in[1,\alpha],i_1<i_2<\cdots<i_{\alpha-j}}P(Q_{i_1}\wedge Q_{i_2}\wedge\cdots\wedge Q_{i_{\alpha-j}})+\cdots+(-1)^{\alpha-1}P(Q_1\wedge Q_2\wedge\cdots\wedge Q_\alpha)). \tag{5}$$

攻击者未获取 i 个特定密钥(i 个密钥不包含在被捕获的 Y 个节点中)的概率可表示为

$$P(Q_1\wedge Q_2\wedge\cdots\wedge Q_i)=\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^Y,$$

其中,X 表示密钥池的规模,α 表示每个节点的私钥数.由此可得,

$$P_{\text{tar}}=\sum_{i=0}^{\alpha}(-1)^iC_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^Y. \tag{6}$$

由索引值的构建原理可知,对于每个传感器节点持有的 α 个私钥,该节点均与其父节点或子节点共享其中的部分密钥.节点在构建索引值时使用的密钥越多,攻击者就越难破坏其隐私性.因此,最坏情况下的隐私泄露发生在叶子节点.假定一个叶子节点有 φ 个祖先,φ 表示聚合树的高度,其取值通常依赖于网络规模、密度及连通度等特性.根据以下公式,可得叶子节点与其祖先共享的密钥数:

$$N_\phi=\sum_{j=0}^{\alpha}(\alpha-j)\left(\sum_{i=j}^{\alpha}(-1)^{i-j}C_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^\phi\right). \tag{7}$$

进而,可得叶子节点索引值失效的概率:

$$p_\varphi=\sum_{j=0}^{\alpha}\left(\sum_{i=0}^{\alpha-j}(-1)^iC_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^Y\right)\times\left(\sum_{i=j}^{\alpha}(-1)^{i-j}C_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^\varphi\right). \tag{8}$$

由于节点采用了数据扰动技术对原始数据进行保护,即 $C_i=2^{l+\text{lb}N}\times t_i+d_i$,所以,即使攻击者得到叶子节点的索引值,也无法直接获知其原始数据.由式(3)可知,攻击者可通过捕获该叶子节点的父节点以进一步获取其原始数据.显然,无任何兄弟节点的叶子节点与其父节点的计算关系最直接,因而最容易发生隐私泄露.假定网络中共有 $m(1\leqslant m<\frac{N}{2})$ 个无兄弟节点的叶子节点,则在攻击者捕获的 Y 个节点中,存在此类节点的父节点的概率为

$$P_m=1-\frac{C_{N-2m}^{Y}}{C_{N-m}^{Y}}, \tag{9}$$

所以,最坏情况下叶子节点隐私泄露的概率为

$$p_{\text{disclosure}}=\sum_{j=0}^{\alpha}\left(\sum_{i=0}^{\alpha-j}(-1)^iC_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^Y\right)\times\left(\sum_{i=j}^{\alpha}(-1)^{i-j}C_\alpha^i\left(\frac{C_{X-i}^{\alpha}}{C_X^{\alpha}}\right)^\varphi\right)\times\left(1-\frac{C_{N-2m}^{Y}}{C_{N-m}^{Y}}\right)\times\left(1-\frac{C_{N-2m}^{Y}}{C_{N-m}^{Y}}\right).$$

4.2 完整性

对于所有的外部节点攻击,每个节点 N_i 在提交数据时均利用与其父节点 N_j 的独立共享密钥 k_{ij} 构造消息认证码 mac_{ij}.如果 N_i 向 N_j 发送的消息被外部攻击者篡改或伪造,N_j 将在验证时发现错误,即 $mac_{ij}\oplus mac_{ji}\neq 0$.同时,节点在提交数据时采用了时间戳技术,如果网络中存在重放攻击,上级节点在利用消息认证码对接收消息进行验证时同样能够发现错误.

5 性能分析

5.1 通信复杂度

在聚合响应过程中,每个节点 N_p 需向父节点 N_q 提交消息 $S_p\parallel IX_{\text{A},p}\parallel mac_{pq}\parallel T$.根据基础密文的构建,$C_i=2^{l+\text{lb}N}\times r_i+d_i$.由 $r_i=\theta$,原始数据 $d_i=l$ 可得,基础密文消息长度为 $\theta+\lfloor\text{lb}N\rfloor+l$.由于 $|IX_{\text{A},p}|=\lambda$,且 θ,l,N,mac_{pq} 和时间戳 T 的长度均为常数,所以节点的通信复杂度为 $O(1)$.

5.2 计算复杂度

和一些典型的聚合算法相比,本文所提算法并未引入额外的消息,而是在聚合结果的基础上增加了 λ 位的索引值作为隐私保护信息.在该协议中,每个节点应用自己的私钥计算其索引值.由于每个节点的私钥数为 α(α 为常数),且 $\alpha<X$,结合基础密文的构建过程可知,节点的计算复杂度为 $O(1)$.

5.3 能量消耗

本方案的能量消耗主要来自于构建非交互式密钥对时昂贵的 ECDH 操作.但是,在本方案中每个传感器节点均和其直接相连的节点构建一对独立共享密钥,且该构建过程仅在系统启动时执行一次.假定采用传感器节点 MICAz[10],且应用 160 b 的椭圆曲线(可取得与 1 024 b 的 RSA 相同的安全级别[11]).由文献[12]可知,在该平台下建立一对非交互式共享密钥所需能量仅为 50.82 mJ.因此,该 ECDH 操作不会给系统造成较重的负担.

表1给出了本方案与部分现有安全聚合协议各项性能指标的对比结果.其中,C表示簇规模.由表1可以看出,在相同的安全级别下,本方案具有更高的能效性.

Table 1　Properties Evaluation of Protocols

表1　协议性能比较

Property	AHE	CPDA	SMART	Our Scheme
Aggregation Type	Hop to Hop	Cluster Head	Hop to Hop	Hop to Hop
Encryption Type	Cluster Head-BS	Node-Cluster Head	Node-Node	Node-Node
Privacy from Eaves	Yes	Yes	Yes	Yes
Privacy from Other Nodes	Yes	Yes	Yes	Yes
Privacy from the BS	No	Yes	Yes	Yes
Data-loss Resilience	Yes	Yes	No	Yes
Intrusion Detection	No	No	No	No
Computation Complexity	$O(1)$	$O(C^2\log C)$	$O(1)$	$O(1)$
Communication Complexity	$O(\log C)$	$O(C)$	$O(C)$	$O(1)$

5.4　仿真实验

在覆盖面积为200 m×200 m的区域内随机部署1 000个节点,对所提算法的隐私性进行实验观察.假设密钥池规模$X=2\,000$,节点传输半径为15 m,妥协节点的比例$\rho=Y/X$,且$\rho\leqslant 30\%$.用$\beta=\alpha/X$表示节点选取私钥数α占密钥总数的比例.

1) 节点隐私泄露概率随α,ρ的变化情况

由图4可见,当妥协节点所占比例ρ较大时,隐私泄露的概率随β的减小而降低.这是因为节点只能使用与其父节点或子节点共享的私钥构建索引值.β的增加将带来2方面的影响:①节点将会使用更多的密钥构建索引值以生成更为复杂的扰动数据,进而增强了抵抗外部攻击者的能力;②如果节点被捕获,攻击者将会获取更多的敏感信息,从而增大隐私数据泄露的风险.从图4可以看出,当妥协节点数超过15%时,负面影响增大,数据隐私泄露的概率将随β的增加而增加;反之,正面影响起主导作用,数据隐私泄露的概率将随β的增大而降低.由此可知,实验结果与式(6)～(9)的理论分析结果保持一致.

2) 抗数据丢失能力

当部分节点数据未被基站成功接收时,便会出现数据丢失情况.为对协议的该特性进行评估,本文采用TAG[13]进行对比.本文所提机制为每个节点的原始数据添加了部分扰动信息,所以可将其封装为一个数据包以对丢包率进行评估,进而获知参与节点的比率.实际上,数据丢失的概率会受很多因素影响,如节点密度、通信量等.本文对此进行了简单分析,且将所有这些复杂因素抽象为一个基本决定因素——每位丢失(错误)率.如果一个消息中出现一个错误位,整个消息都要被丢弃.因此,本文定义

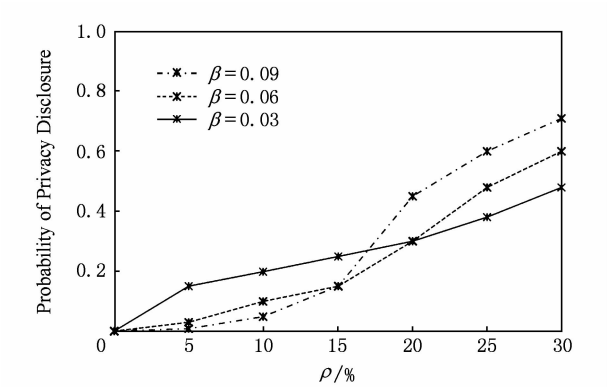


Fig. 4　Probability of privacy disclosure with different compromised nodes.

图4　隐私数据泄露与妥协节点数的关系

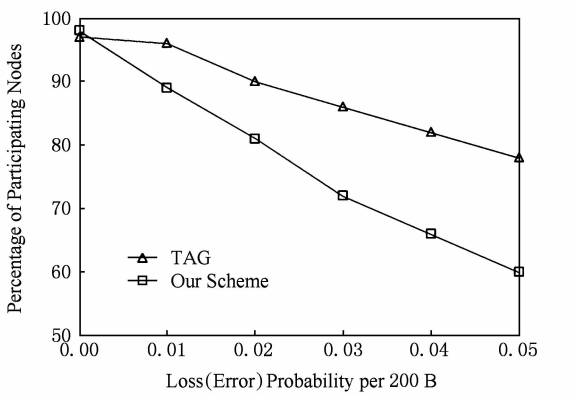


Fig. 5　Percentage of nodes actually participating in data aggregation.

图5　实际参与数据聚合的节点比例

位丢失(错误)率,并就实际参与聚合的节点数情况与 TAG 进行对比.由图 5 可见,当每位丢失率较低时,本文所提方案与 TAG(无任何隐私保护策略)中实际参与聚合的节点数相当.另外,因 TAG 中消息的规模约为 200 B,为便于比较,图 5 中 X 轴以每 200 B 丢失率为单位进行对比.

6 结 论

本文对 WSNs 中聚合数据的隐私保护问题进行了分析,提出一种能量有效的隐私保护数据聚合方案.通过进行 2 次不同形式的数据扰动分别实现单个节点数据对基站及邻居节点的隐私保护.安全及性能分析表明,所提方案可在不过多消耗节点能量的前提下保证节点的安全性,且具有较好的抗数据丢失能力,安全性及能效性均优于现有方案.

参 考 文 献

[1] Ramesh R, Varshney P K. Data aggregation techniques in sensor networks: A survey [J]. Communications Surveys & Tutorials, 2006, 8(4): 48-63

[2] Yick J, Mukherjee B, Ghosal D. Wireless sensor network survey [J]. Computer Networks, 2008, 52(12): 2292-2330

[3] He Weibo, Nguyen H. PDA: Privacy-preserving data aggregation in wireless sensor networks [C] //Proc of the 26th IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2007: 2045-2053

[4] Castelluccia C, Mykletun E, Tsudik G. Efficient aggregation of encrypted data in wireless sensor networks [C] //Proc of the 2nd Annual Int Conf on Mobile and Ubiquitous Systems in Computing, Networking and Services. Piscataway, NJ: IEEE, 2005: 109-117

[5] Yang Geng, Wang Anqi, Chen Zhengyu, et al. An energy-saving privacy-preserving data aggregation algorithm [J]. Chinese Journal of Computers, 2011, 34(5): 792-800 (in Chinese)
(杨庚, 王安琪, 陈正宇, 等. 一种低能耗的数据融合隐私保护算法[J]. 计算机学报, 2011, 34(5): 792-800)

[6] Bista R, Kim D, Chang J. A new private data aggregation scheme for wireless sensor networks [C] //Proc of the 10th Int Conf on Computer and Information Technology. Piscataway, NJ: IEEE, 2010: 273-280

[7] Mao Wenbo. Modern Cryptography: Theory and Practice [M]. Upper Saddle River, NJ: Prentice Hall, 2003: 118-160

[8] Colin B, Mao Wenbo, Kenneth G P. Key agreement using statically keyed authenticators [C] //Proc of the 2nd Int Conf on Applied Cryptography and Network Security (ACNS'04). Berlin: Springer, 2004: 248-262

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

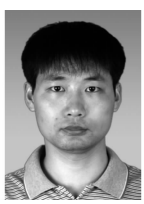
[10] Lu Rongxing, Lin Xiaodong, Zhu Haojin, et al. BECAN: A bandwidth-efficient cooperative authentication scheme for filtering injected false data in wireless sensor networks [J]. IEEE Trans on Parallel and Distributed Systems, 2012, 23(1): 32-43

[11] Mao Wenbo. Modern Cryptography: Theory and Practice [M]. Upper Saddle River, NJ: Prentice Hall, 2003: 212-236

[12] Liu An, Ning Ping. TinyECC: A configurable library for elliptic curve cryptography in wireless sensor networks [C] //Proc of the 7th Int Conf on Information Processing in Sensor Networks (IPSN'08). New York: ACM, 2008: 245-256

[13] Madden S, Franklin M J, Hellerstein J M, et al. TAG: A tiny aggregation service for ad-hoc sensor networks [C] //Proc of the 5th Symp on Operating Systems Design and Implementation. Berkeley, CA: USENIX Association, 2002: 131-146

Fu Shuai, born in 1986. PhD and assistant professor. Student member of China Computer Federation. Her main research interests include wireless network security, wireless sensor network and data aggregation.



Jiang Qi, born in 1983. PhD and associate professor. Member of China Computer Federation. His main research interests include protocol security analysis and wireless network security (jiangqixdu@gmail.com).

Ma Jianfeng, born in 1963. Professor and PhD supervisor at the School of Computer Science and Technology, Xidian University. Senior member of China Computer Federation. His main research interests include distributed systems, wireless and mobile computing systems, and network and information security (jfma@mail.xidian.edu.cn).