研究与开发

面向物联网轻量级隐私保护的真值发现机制

何英杰，李启伟，孙涵，郜迪，董剑峰，杨书华
（中国电子科技集团公司第二十二研究所，河南 新乡 453000）

摘　要：针对云雾融合物联网系统中感知数据的质量差异和隐私泄露问题，提出一种基于流加密模式的隐私保护真值发现机制。首先，云服务器将洗牌算法和流加密算法相结合，实现匿名更新真值；雾服务器匿名更新权重，防止恶意攻击者与云雾服务器的共谋攻击所造成的设备隐私泄露。其次，雾服务器基于 Softmax 函数计算设备权重，以减小计算真值的误差。最后，理论分析证明该机制能够保护设备隐私。实验结果表明，该机制误差率低，并在时间开销上优于已有方法，是一种适用于规模化物联网高效的真值发现机制。

关键词：物联网；轻量级；隐私保护；真值发现
中图分类号：TN918
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2021100

A lightweight privacy-preserving truth discovery mechanism for IoT

HE Yingjie, LI Qiwei, SUN Han, GAO Di, DONG Jianfeng, YANG Shuhua
The 22nd Research Institute of China Electronics Technology Group Corporation, Xinxiang 453000, China

Abstract: In order to solve diverse quality and privacy leakage of perceived data in fog-cloud integrated internet of things (IoT), a streaming encryption-based privacy-preserving truth discovery mechanism for IoT was proposed. Firstly, by utilizing shuffling and streaming encryption algorithms, the ground truths and the weights were anonymously updated on the cloud server and fog server, respectively, so that the collusion attacks between malicious attackers and cloud or fog servers could be resisted to defend against privacy leakage of IoT devices. Secondly, by adopting the Softmax function, the device weights were calculated on fog server, which reduces the error rate for calculating the ground truths. Finally, the theoretical analysis proved that the mechanism could protect privacy of the devices. And, the experimental results demonstrate that the proposed mechanism is an effective privacy-preserving trust discovery mechanism for large-scale IoT devices, which can outperform existing ones in computing efficiency.

Key words: IoT, lightweight, privacy-preserving, truth discovery

1　引言

物联网各种各样的应用将各个终端设备数据聚合到数据中心，通过对设备数据的统计分析来为消费者提供优质的服务[1]，例如智慧家居[2]、智慧交通[3]、智慧医疗[4]等，这些应用提高了人们的

收稿日期：2020−09−24；修回日期：2021−05−10

生活质量并产生了巨大的社会经济效益[5]。

物联网应用利用聚合数据产生价值，物联网数据质量越高，数据价值也就越大。但是，不同的物联网设备采集的数据质量具有差异性[6]。首先，引起物联网数据质量差异性的主要原因是数据的不确定性，例如，数据聚合的环境动态变化可使数据产生不定因素（错误、噪声、分布状态等）[7]，从而影响原始数据质量；其次，物联网系统可能存在恶意节点，恶意节点产生异常数据，若这些质量较低的异常信息聚合到数据中心，则会影响应用的实际效果；此外，物联网系统中存在大量收集用户敏感数据的设备，这些设备常常成为恶意攻击者的目标，例如，物联网中的共谋攻击[8]、扰动攻击[9]或者选择明文攻击（chosen-plaintext attack，CPA）[8]等都会引起数据隐私的泄露问题。因此，提高聚合数据的质量以及保护数据隐私对物联网应用而言极其重要。

为了应对提高聚合质量的挑战，真相发现最近已得到广泛研究[10]，旨在从不可靠的数据中发现真值。真相发现方法的基本原则是：若感知的数据更接近汇总结果，则将为其分配较高的权重；而若数据具有较高的权重，则将在聚合过程中，其数据获得更多统计。

尽管真相发现方法已极大地改善系统中聚合的准确性，但仍无法解决参与设备的隐私问题。在 IoT 应用中，参与者通常不愿提供其真实的感知数据，这些感知数据中推断得到参与者的隐私信息。例如众智感知系统[10]中能够通过汇总从大量设备那里收集的数据来解决一些难题，但是，平台系统能够从设备的收集的回答中推断得到每个参与设备所属人员的隐私信息；此外，聚合健康数据[11]（例如治疗结果）能够更好地评估新药或医疗设备的效果，但可能会泄露参与患者的隐私；地理标记[12]能够通过汇总参与者的报告来提供对特定对象（例如垃圾、坑洼、自动外部去纤颤器等）的准确、及时的定位，但存在泄露参与者敏感位置信息的风险。

目前，已有研究方法关注解决聚合数据的质量和隐私泄露问题。Xu 等人[13]提出了一种面向移动群智感知系统的高效保护隐私的真值发现（efficient and privacy-preserving truth discovery，EPTD）方法，但用户组之间生成大量共享密钥而导致其能耗成本开销较大；Tang 等人[14]提出了一种面向群智感知系统的非交互隐私保护真值发现（privacy-preserving truth discovery，PPDT）算法，但算法的加速器只针对特定问题场景；Zhang 等人[15]提出了一种面向移动群智感知系统的可靠隐私保护真值发现算法，但算法的计算成本较高；Mao 等人[16]提出了基于两个非协同的云服务器的轻量级真值发现框架分别为 L-PPDT（lightweight privacy preserving truth discovery framework）及 L^2-PPDT（more lightweight privacy preserving truth discovery framework）。该方案中，用户能够选择不同的隐私加密级别，但是，可信度高的用户的数据通常具有更高价值，其方案并不能实现差异化的隐私保护级别。

针对上述问题，本文提出了一种面向物联网的轻量级真值发现（lightweight privacy-preserving truth discovery for IoT，IoT-LPTD）机制，该机制提高了聚合数据质量和保护了数据隐私，主要贡献如下。

（1）物联网的云服务器基于洗牌算法和流式加密实现了匿名计算真值。在云服务器计算真值的过程中，即使恶意终端设备和云服务器发起共谋攻击，诚实的终端设备的隐私也不会被泄露。

（2）雾服务器和可信服务器基于密文协同计算设备的可信权重，防止了雾服务器泄露设备隐私。同时，雾服务器采取了 Softmax 函数计算设备的权重，减少了计算真值的误差率。

（3）理论分析证明了该机制具有保护设备隐私的特点。实验结果表明，该机制误差率低，并在时间开销上优于已有方法。

2 预备知识

2.1 双线性映射

设 q 是一个大素数，G_1 和 G_2 分别为两个 q 阶的循环乘法群。G_1 到 G_2 的双线性映射为 $e: G_1 \times G_2 \to G_3$，其满足下列性质[12]。

（1）双线性：对于任意的 $P,Q,R \in e(P,Q)^{ab}$ 或者 $e(P+R,Q)=e(P,Q)\cdot e(R,Q)$，$e(P,Q+R)=e(P,Q)\cdot e(P,R)$，那么该映射称为双线性映射。

（2）非退化性：如果 P 是 G_1 的生成元，那么 $e(P,P)$ 就是 G_2 的生成元。

（3）可计算性：对于任意的 P、Q，在有效的时间内能够计算得到 $e(P,Q)$。

2.2 真值发现算法

真值发现算法[13]主要分为两个步骤：权重更新和真值更新。

（1）权重更新：由于每个设备感知数据 x 和真值 x_t 存在差别，根据感知数据的差异，对每个设备定义一个可信的权重 $w=f(d(x,x_t))$，其中，$f(\cdot)$ 是一个单调递减函数，$d(x,x_t)$ 用来测量感知数据和预测真值之间的差异。

（2）真值更新：根据每个设备权重值计算得到当前的真值 x_t：

$$x_t=\frac{\sum_{m=1}^{M} w_m x_m}{\sum_{m=1}^{M} w_m} \tag{1}$$

其中，M 是设备的数量，x_m 是第 m 设备感知数据值。

3 系统模型

3.1 IoT-LPTD 模型

IoT-LPDT 模型示意图如图 1 所示，IoT-LPTD 由可信服务器（trust authority server，TA）、云服务器（cloud server，CS）、雾服务器（fog server，FS）、设备节点（device node，DN）组成，其建模为四元组（TA,FS,CS,DN）。在 IoT-LPTD 模型中，首先，TA 为 CS、FS、DN 分发签名密钥对 $<sk,pk>$ 和传输密钥（K），并维护一个节点权重列表（L_T），同时 CS 在执行真值算法之前，TA 将发送给 CS 一个扰动的权重列表（W_t）；其次，CS 在 t 时刻请求 DN 中的设备节点（d_i）采集不同对象的属性（x_{o_i}），当 d_i 采集到数据集合 X_O 之后，执行加密数据算法和签名算法，并将密文传送到 FS 中的雾服务器（f_i），f_i 执行密文聚合算法并把聚合密文传送至云服务器（c_i），c_i 匿名地计算真值并加密得到真值集合 X_O^*，FS 接收加密的真值集合后，执行计算可信权重算法并把新的可信权重列表（W_{t+1}）传送给 TA；最后，TA 更新 L_T，其中，IoT-LPTD 模型中所用的数学符号见表 1。

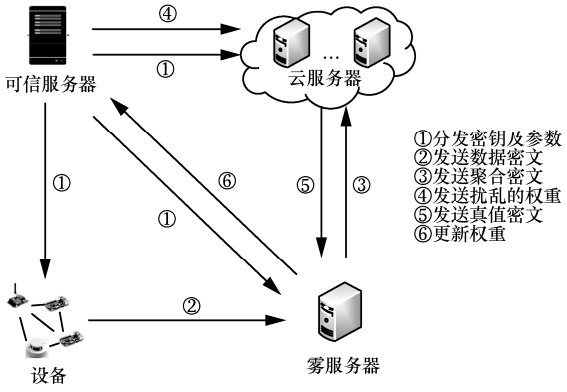

图 1 IoT-LPDT 模型示意图

表 1 符号说明

符号名称	说明
t	时间戳
TA	可信服务器
L_T	权重列表
K	传输密钥集
$CS:\{1,\cdots,c_i\}$	云服务器
$FS:\{1,\cdots,f_i\}$	雾服务器
$DN:\{1,\cdots,d_i\}$	设备节点
$O:\{1,\cdots,o_i\}$	数据对象
$X_O:\{x_1,\cdots,x_{o_i}\}$	数据对象的属性
$CT:\{ct_1,\cdots,ct_i\}$	数据密文
$X_O^*:\{x_1^*,\cdots,x_{o_i}^*\}$	数据真值
$CX^*:\{cx_1^*,\cdots,cx_{o_i}^*\}$	真值密文
$W_t:\{w_1,\cdots,w_n\}$	设备可信权重

3.2　威胁模型

假设物联网系统模型是半诚实的，即在威胁模型中的参与者将会严格遵守设计的 IoT-LPTD 机制；但是，云服务器或雾服务器能够通过收集感知数据去推断每个设备的敏感信息。每个设备能够尝试获取其他设备的敏感数据，进行推断或学习其他设备隐私。

4　IoT-LPTD 实现

IoT-LPTD 主要由 4 个阶段组成：系统建立、感知数据加密、真值匿名计算以及权重匿名更新。

4.1　系统建立

4.1.1　系统初始化

TA 设置一个初始化的安全参数 κ，由安全参数和大素数生成算法生成一个较大的素数 q。TA 构建两个循环群 G_1、G_2，两个循环群的阶数为最大素数 q。设 g_1、g_2 分别为 G_1、G_2 的生成元，其双线性映射为 $e:G_1\times G_2\to G_T$。系统构造全域且单向的哈希函数 $H:\{1,0\}^*\to G_1$ 和以密钥 k 为种子的伪随机函数 $f_k(x)$。

$$f_k(x):H_{l,N+l_t}=\{f_k:\{1,0\}^{l_t+N}\to\{1,0\}^{l}\}_{k\in\{1,0\}^{l}} \quad (2)$$

其中，l、N、l_t 分别表示密文长度、明文允许的最大长度以及时间戳长度。故 TA 生成公共的系统参数为 $\{k,q,G_1,G_2,G_T,g_1,g_2,e,H,f_k(x)\}$。

4.1.2　密钥生成

TA 根据公共参数为 DN、FS、CS 生成签名密钥对 $<\mathrm{sk},\mathrm{pk}>$，传输密钥（$K$），解密密钥 MASK。

（1）DN：TA 为 d_i 从 Z_p 群中选择一个随机数 s，令其作为 d_i 的签名私钥，即 $\mathrm{sk}_{d_i}=s$，与此同时计算与之对应的签名公钥：$\mathrm{pk}_{d_i}=g^s$，然后 TA 根据 d_i 的真实 ID_{d_i} 生成一个伪身份 $\mathrm{PID}_i^d=H(\mathrm{ID}_{d_i}\,|\,\mathrm{sk}_{d_i})$。

（2）FS、CS：TA 使用与设备节点相同的原理 TA 为 f_i 生成签名密钥对 $<\mathrm{sk}_{f_i},\mathrm{pk}_{f_i}>$ 和伪身份 PID_i^f，同时为 c_i 生成签名密钥对 $<\mathrm{sk}_{c_i},\mathrm{pk}_{c_i}>$ 和伪身份 PID_i^c。

（3）FS、CS 的 K 构建：TA 为 d_i 从 Z_p 中挑选两个随机数 α_{d_i}、β_{d_i} 作为数据加密的密钥 $k_{d_i}=<\alpha_{d_i},\beta_{d_i}>$，$\alpha_{d_i}$、$\beta_{d_i}$ 分别存储到 TA 的密钥列表 K_{f_i}、K_{c_i} 中。

（4）解密密钥生成：TA 在时刻 t 根据 K_{f_i}、K_{c_i} 及式（2）分别为 f_i、c_i 计算聚合密钥 MASK_{f_i}、MASK_{c_i}：

$$\mathrm{MASK}_{c_i}=\otimes_{i=1}^{n_{c_i}}f_{\beta_i}(t\,|\,1)\,|\otimes_{i=1}^{n_{c_i}}f_{\beta_i}(t\,|\,2)\,|\cdots|\otimes_{i=1}^{n_{c_i}}f_{\beta_i}(t\,|\,n_{f_j}) \quad (2a)$$

$$\mathrm{MASK}_{f_i}=\otimes_{i=1}^{n_{f_i}}f_{\alpha_i}(t\,|\,1)\,|\otimes_{i=1}^{n_{f_i}}f_{\beta_i}(t\,|\,2)\,|\cdots|\otimes_{i=1}^{n_{f_i}}f_{\alpha_i}(t\,|\,n_{f_j}) \quad (2b)$$

其中，n_{f_i}、n_{c_i} 分别表示 f_i 连接设备数量、c_i 连接雾服务器数量，$\otimes_{i=1}^{n}y=y_1\otimes\cdots\otimes y_n$。

4.1.3　可信权重列表构建

可信权重列表 L_T 见表 2，TA 构建一个可信权重列表以追踪 L_T 恶意节点和匿名实现设备的数据可信权重的更新，其中，设备的初始化权重为 $w_i=1/\mathrm{len}(L_\mathrm{T})$。

TA 通过洗牌算法构建一个可逆函数 $I(x)$，利用 $I(x)$ 的性质为 d_i 的随机生成一个数据加密的位置 j。其中，$I(x)$ 函数满足在不同周期内，相同的 i 输出的位置 j 不同：

$$j=I(i)\Leftrightarrow I^{-1}(j)=i \quad (3)$$

表 2　可信权重列表 L_T

i	伪身份 ID	签名公钥	可信权重	对应雾节点
1	PID_1^d	g^{s_1}	w_1	PID_1^f
2	PID_2^d	g^{s_2}	w_2	PID_2^f
3	PID_3^d	g^{s_3}	w_3	PID_3^f
…	…	…	…	…

4.2　感知数据加密

当 d_i 采集到数据 X_o 后，设备 d_i 通过加密算法和签名算法保护敏感数据的隐私。具体过程如下。

（1）在 t 时刻下，d_i 向 TA 请求一个数据有效

索引位置 $I(i)$，并对感知数据明文加密：

$$\mathrm{CT}_{i,I(i)} = X_o \otimes f_{\alpha_i}(t \mid I(i)) \otimes f_{\beta_i}(t \mid I(i)) \quad (4)$$

（2）令 $j = 1,\cdots,n_{f_i}$ 且 $j \neq I(i)$，计算第 j 处的密文：

$$\mathrm{CT}_{i,j} = f_{\alpha_i}(t \mid j) \otimes f_{\beta_i}(t \mid j) \quad (5)$$

（3）对密文集合进行聚合：$\mathrm{CT}_i = \mathrm{CT}_{i,1} \mid \mathrm{CT}_{i,2} \mid \cdots \mid \mathrm{CT}_{i,n_{fi}}$。

（4） 密文签名:

$$\delta_i = H_{sk_{d_i}}(\mathrm{CT}_i \mid \mathrm{PID}_i^d) \quad (6)$$

当 d_i 完成上述步骤之后，把 $\mathrm{data}_i =< \mathrm{PID}_i^d, \delta_i, \mathrm{CT}_i >$ 发给所连接的雾节点 f_i。

4.3 真值匿名计算

f_i 收集所连接设备的数据 data_i 后，向可信服务器请求 t 时刻的 MASK_{f_i}，并执行已经初始化的验证聚合签名算法和聚合算法。f_i 把聚合密文上传到云服务器，云服务器得到聚合密文进行解密，然后计算真值。此时，云服务器即使得到设备的原始数据，也无法推断具体设备上传的原始数据，算法详细步骤如下。

（1）聚合签名验证。

$$\delta_{\mathrm{agg}} = \prod_{i=1}^{n_{fi}} \delta_i \quad (7)$$

令 $i = 1,\cdots,n_{f_i}$，计算：

$$\begin{cases} \delta_i^{'} = H_i(\mathrm{CT}_i, \mid \mathrm{PID}_i^d) \\ e(\delta_{\mathrm{agg}}, g_2) \stackrel{?}{=} \prod_{i=1}^{n_{fi}} e(\delta_i^{'}, \mathrm{pk}_{d_i}) \end{cases} \quad (8)$$

（2）若聚合签名不合法，则输出错误消息并把 PID_i^d 发送给可信服务器，可信服务器在可信的权重列表中找出非法节点。若聚合签名合法，则执行雾节点的密文数据聚合算法：

$$\begin{cases} \mathrm{CT}_{\mathrm{agg}} = \mathrm{CT}_1 \otimes \mathrm{CT}_2 \otimes \cdots \otimes \mathrm{CT}_{n_{fi}} \\ \mathrm{CT}_{\mathrm{agg}}^{'} = \mathrm{CT}_{\mathrm{agg}} \otimes \mathrm{MASK}_{f_i} \end{cases} \quad (9)$$

（3）对密文 $\mathrm{CT}_{\mathrm{agg}}^{'}$ 进行签名：

$$\delta_{f_i} = H_{\mathrm{sk}_{fi}}(\mathrm{CT}_{\mathrm{agg}}^{'} \mid \mathrm{PID}_i^f) \quad (10)$$

雾节点 f_i 发送 $\mathrm{data}_{f_i} =< \mathrm{PID}_i^f, \delta_{f_i}, \mathrm{CT}_{\mathrm{agg}}^{'} >$ 至 c_i，当 c_i 接收 f_i 的数据 data_{f_i} 后，首先向可信服务器请求 t 时刻的 MASK_{c_i}，然后执行数字签名的验证算法验证数据合法性，如果合法进行解密数据并计算真值。算法详细步骤如下。

（1） 数字签名验证：

$$e(H(\mathrm{CT}_{\mathrm{agg}}^{'} \mid \mathrm{PID}_i^f), g_2) \stackrel{?}{=} \prod_{i=1}^{n_{fi}} e(\delta_{f_i}, \mathrm{pk}_{f_i}) \quad (12)$$

（2）根据式（13）解密密文得到明文 $X = \{X_1, X_2, \cdots, X_{n_{ci}}\}$：

$$X = \mathrm{CT}_{\mathrm{agg}}^{'} \otimes \mathrm{MASK}_{c_i} \quad (13)$$

（3）根据云服务器在有效时间段内收到可信服务器下发的一次性权重列表 W_t 和式（1）计算 X_t：

$$X_t = \frac{\sum_{w_i \in W_t, x_i \in X}^{n_{ci}} x_i w_i}{\sum_{w_i \in W_t}^{n_{ci}} w_i} \quad (14)$$

（4）加密 X_t 及签名:

$$\begin{cases} \mathrm{CX}_t^* = X_t \otimes \mathrm{MASK}_{c_i} \\ \delta_{c_i} = H_{\mathrm{sk}_{c_i}}(\mathrm{CX}_t^* \mid \mathrm{PID}_i^c) \end{cases} \quad (15)$$

c_i 把数据 $\mathrm{data}_{c_i} =< \mathrm{PID}_i^c, \delta_{c_i}, \mathrm{CX}_t^* >$ 发送至对应的 f_i。

4.4 权重匿名更新

f_i 接收云服务器数据 data_{c_i} 后，验证数据合法性并在密文上计算设备权重。算法详细步骤如下。

（1）验证签名：

$$e(H(\mathrm{CX}_t^* \mid \mathrm{PID}_i^c), g_2) \stackrel{?}{=} e(\delta_{c_i}, \mathrm{pk}_{c_i}) \quad (16)$$

（2） 若签名合法，则 f_i 计算感知数据与真值的差异度：

$$\begin{cases} \mathrm{DIF} = \{\mathrm{DI}_1, \cdots, \mathrm{DI}_{n_{fi}}\} \\ \mathrm{DI}_i = \{\mathrm{di}_1, \cdots, \mathrm{di}_{n_{fi}}\}, \mathrm{di} = x_o^* \otimes x_o \\ \mathrm{DIF} = \mathrm{CT}_t^* \otimes \mathrm{CT}_{\mathrm{agg}}^{'} \end{cases} \quad (17)$$

（3）基于 Softmax 函数计算设备的权重 W_{t+1}：

$$W_{t+1}=\{w_1,w_2,\cdots,w_{n_{f_i}}\}=\left\{\frac{\sum_{j=1}^{n=\text{len}(\text{DI}_i)}\text{e}^{di_j}}{\sum_{n=\text{len}(\text{DIF})}\sum_{n=\text{len}(DI_i)}\text{e}^{di_j}}\right\} \quad (18)$$

f_i 计算得到的 $<W_{t+1},\text{PID}_i^f>$ 发送给 TA。TA 执行更新可信权重列表操作，具体操作如下。

根据 PID_i^f 对应的雾节点连接的设备的可信权重，然后更新：

$$L_{\text{T}}\{W\}=\{w_i\}=W_{t+1}\{w_{I^{-1}(i)}\},i=1,\cdots,\text{len}(W_{t+1}) \quad (19)$$

5　IoT-LPTD 分析

5.1　安全分析

在第 3.2 节所述的威胁模型下，从设备、雾服务器、服务器的匿名计算以及加密模式的安全性两个方面对 IoT-LPDT 进行安全分析。

5.1.1　匿名计算安全分析

定义　设备感知数据 $M=\{m_1,\cdots,m_n\}^N\in\{M^N\}$，其中 $M\in\{0,1\}^l$，每个设备 d_i 执行感知数据加密算法加密 m_i 生成密文 ct_i，FS 收集 $\text{CT}:\{\text{ct}_1,\cdots,\text{ct}_i\}$ 并执行数据验证和聚合算法生成 CT'_{agg}，CS 最后从 FS 收集 CT'_{agg}。

对于 DN，由于每个设备不能获取 FS 和 CS 的私钥，设备不能获得其他设备的原始数据。所以设备不能破坏整个协议匿名计算真值的特性。

对于 FS，虽然能从所有设备收集 $\text{CT}:\{\text{ct}_1,\cdots,\text{ct}_i\}$，但是由于它缺少 CS 的解密密钥，没有能力恢复 DN 的任何设备的感知数据。其次，雾服务器使用设备收集的密文 $\{\text{ct}_1,\cdots,\text{ct}_n\}$ 和基于云服务器密钥加密后的真值密文进行异或运算，然后根据式（18）计算得到各个设备的权重，并且根据假名机制记录在可信服务器 TA。值得注意的是，设备的权重是基于密文计算的。因为 FS 也无法获取 CS 的私密钥和时间戳等加密参数，FS 根据异或之差反推得到设备的原始数据，所以 FS 不能够破坏整个协议匿名计算真值的特性。

对于 CS，虽然能够恢复所有设备的感知数据，但是，CS 无法区分具体某一个设备的隐私数据，即：如果一个攻击者仅知道数据来源于 k 个设备数据源，但是在多项式时间内，无法区分数据来源于具体某一个设备，则说明该协议具有匿名性[16]。因此，如果 IoT-LPDT 在同一周期下，某两个设备的密文发生位置交换，CS 无法区分交换前后密文的差异性，则证明 IoT-LPDT 协议具有匿名性。然而，CS 仅能够获取的 CS 密文为：

$$V_1=\text{C}'_{\text{agg}}=m_{I^{-1}(1)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(1|t)m_{I^{-1}(1)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(1|t)|\cdots|$$
$$m_{I^{-1}(a)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(a|t)|\cdots|m_{I^{-1}(b)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(b|t)|\cdots|$$
$$m_{I^{-1}(n)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(n|t)$$

令 $c_{0,1}=m_{I^{-1}(j)}\overset{n_{ck}}{\underset{i=1}{\otimes}}f_{\beta_i}(j|t)$，$j\in[1,n]$，则 $V_1=c_{0,1}|c_{0,2}|\cdots|c_{0,n}$。

假设设备 d_a 和 d_b 的交换的数据为 m_a、m_b，其中 $1\leqslant a<b\leqslant n$，那么所有设备数据为 $M'=\{m_1,\cdots,m_{a-1},m_b,m_{a+1},\cdots,m_{b-1},m_a,m_{b+1},\cdots,m_n\}$，则密文变为 $V_2=c_{0,1}|\cdots|c_{0,a-1}|c_{0,b}|c_{0,a+1}|\cdots|c_{0,b-1}|c_{0,a}|c_{0,b+1}|\cdots|c_{0,n}$。

因此证明协议具有匿名性，必须证明对于任意的 a、b，$V_1\equiv V_2$。假设存在一个模拟设备 d_s 执行该协议，首先生成一个 $\{1,\cdots,n\}$ 的随机排列组合 $\{I^{-1}(1),\cdots,I^{-1}(n)\}$，然后 d_s 得到 M'' 为：

$$M''=\{m_{I^{-1}(0)},\cdots,m_{I^{-1}(n)}\}$$

最后 d_s 输出的密文为：$V_3=c_{0,I^{-1}(0)}|c_{0,I^{-1}(2)}|\cdots|c_{0,I^{-1}(n)}$。

因为在多项式时间内，M 和 M'' 不能被区分，所以，V_1 和 V_3 也不能区分。因此，$V\equiv V''$，同理，在多项式内，M' 和 M'' 不能被区分。因此，$V_1\equiv V_2$。

因为 $V_1\equiv V_2$，所以 IoT-LPTD 协议满足 K 匿

名性。

最后，IoT-LPTD 即使根据式（14）计算得到真值 X_t，但是整个计算过程，CS 无法区分设备敏感数据。

综上所述：IoT-LPTD 具有匿名计算真值的性质。

5.1.2 加密模式安全分析

在云服务器计算真值过程中，设备端通过异或操作把数据密文聚合到云服务器。参考文献[17]的对称加密模式 $c=<r, f_{k_1}(r)\otimes m>$ 是满足 CPA 安全要求的，其中，r 是随机数，而 IoT-LPTD 协议的加密模式 $c=<r, f_{k_1}(r)\otimes f_{k_2}(r)\otimes m>$ 也是满足 CPA 安全要求的。

证明 参考文献[17]已经证明 $c=<r, f_{k_1}(r)\otimes m>$ 满足 CPA 安全要求。令 $k_1\neq k_2$，$m'=f_{k_2}(r)\otimes m$，则 $c=<r, f_{k_1}(r)\otimes m'>$ 满足 CPA 安全要求。因此，$c=<r, f_{k_1}(r)\otimes f_{k_2}(r)\otimes m>$ 也满足 CPA 安全要求。

综上所述：IoT-LPTD 也满足 CPA 安全要求。

5.2 性能分析

仿真实验分析了 IoT-LPTD 的性能。其实验设置如下。

多个设备在 t 时刻多次采集不同数量的数据，其中，数据的取值服从于置信区间为 0.95、均值为 0、方差为 10.5 的正态分布，若数据取值不在置信区间内，则认为该数据为异常值。

其次，设置加密的初始化参数为 $\{k=160, f_k(x):\text{HMAC\_SHA512}\}$。

5.2.1 时间开销分析

系统时间开销见表 3，其表明在时间开销方面，IoT-LPTD 优于已有方法。

与已有方法相比，IoT-LPTD 的时间开销最低。主要原因分析如下：在加解密方面只有异或操作运算，减少了加解密的时间消耗成本。相比上述 4 种方案，它采取云端计算真值更新和权重更新两个步骤，而 IoT-LPTD 则在云端只计算真值，而雾节点协助云端计算真值，从而减少了云端的计算成本。

表 3 系统时间开销

方法	数据对象数目		
	4	8	10
L-PPTD[16]	0.125 s	0.2 s	0.275 s
L^2-PPTD[16]	0.175 s	0.275 s	0.350 s
EPTD[13]	0.115 s	0.151 s	0.252 s
DPoR[18]	0.360 s	0.378 s	0.390 s
IoT-LPTD	0.065 s	0.075 s	0.076 s

5.2.2 误差率分析

为评估 IoT-LPTD 可用性，采用参考文献[15]中的平均误差率作为评价指标，平均误差率如式(21)所示：

$$\text{err\_rate}=\frac{|x_t'-x_t^*|}{nx_t^*} \tag{21}$$

其中，x_t' 为云服务器计算出的真值，x_t^* 为真实值，n 表示采集的设备。

采集不同数据对象数目的误差示意图如图 2 所示，实验结果表明，IoT-LPTD 具有高可用性。分析原因如下：与传统的方法采取式（14）作为计算权重的方法相比，IoT-LPTD 采取式（18）计算权重，其目的是使越靠近真值的感知数据的比重越大，从而使计算得到的真值误差率越来越小。

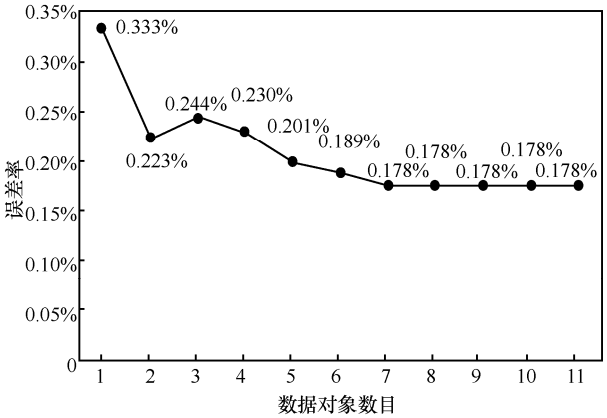

图 2 采集不同数据对象数目的误差示意图

6　结束语

本文针对物联网中终端设备采集感知数据质量参差不齐、聚合过程中泄露隐私数据等问题提出一种面向物联网轻量级隐私保护的真值发现机制。该机制实现了匿名计算真值和基于密文计算权重的功能，抵御了推断或泄露设备隐私的恶意攻击。最后，分析了该机制的安全性。实验表明该机制具有时间开销低、可用性高的特点，是一种适应于规模化物联网高效的真值发现机制。

参考文献：

[1] GAO X, LIU C C. The design of data service system in the IoT resource access and intelligent processing platform[C]// Proceedings of 2016 International Conference on Wireless Communication and Network Engineering (WCNE2016). [S.l.:s.n.], 2016: 393-396.

[2] LEE J, HAN J, CHA J. A study on SSDP protocol based IoT / IoL device discovery algorithm for energy harvesting interworking smart home[J]. International Journal of Internet, Broadcasting and Communication : IJIBC, 2018(10).

[3] CAO X, LI Y Q. Data collection and network architecture analysis in Internet of vehicles based on NB-IoT[C]//Proceedings of 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). Piscataway: IEEE Press, 2018: 157-160.

[4] TANG W J, REN J, DENG K, et al. Secure data aggregation of lightweight E-healthcare IoT devices with fair incentives[J]. IEEE Internet of Things Journal, 2019, 6(5): 8714-8726.

[5] CHOI H S, RHEE W S. IoT-based user-driven service modeling environment for a smart space management system[J]. Sensors (Basel, Switzerland), 2014, 14(11): 22039-22064.

[6] KARKOUCH A, MOUSANNIF H, AL MOATASSIME H, et al. Data quality in Internet of Things: a state-of-the-art survey[J]. Journal of Network and Computer Applications, 2016(73): 57-81.

[7] DHINGRA S, MADDA R B, GANDOMI A H, et al. Internet of Things mobile–air pollution monitoring system (IoT-mobair)[J]. IEEE Internet of Things Journal, 2019, 6(3): 5577-5584.

[8] 谢丽霞, 魏瑞炘. 一种面向物联网节点的综合信任度评估模型[J].西安电子科技大学学报, 2019, 46(4): 58-65.
XIE L X, WEI R X. IoT Node trust comprehensive evaluation model[J]. Journal of Xidian University, 2019, 46(4): 58-65.

[9] 钱亚冠, 关晓惠, 吴淑慧, 等. 一种基于贪心算法的 SVM 扰动攻击方法[J]. 电信科学, 2019, 35(01): 81-89.
QIAN Y G, GUAN X H, WU S H, et al. A SVM disturbance attack method based on greedy algorithm[J]. Telecommunications Science, 2019, 35(01): 81-89.

[10] MIAO C L, JIANG W J, SU L, et al. Cloud-enabled privacy-preserving truth discovery in crowd sensing systems[C]//Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2015:183-196.

[11] MARSHALL J, WANG D. Towards emotional-aware truth discovery in social sensing applications[C]//Proceedings of 2016 IEEE International Conference on Smart Computing (SMARTCOMP). Piscataway: IEEE Press, 2016: 1-8.

[12] LI Z Q, YANG S, WU F, et al. Holmes: tackling data sparsity for truth discovery in location-aware mobile crowdsensing[C]//Proceedings of 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). Piscataway: IEEE Press, 2018: 424-432.

[13] XU G W, LI H W, LIU S, et al. Efficient and privacy-preserving truth discovery in mobile crowd sensing systems[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3854-3865.

[14] TANG X T, WANG C, YUAN X L, et al. Non-interactive privacy-preserving truth discovery in crowd sensing applications[C]//Proceedings of IEEE INFOCOM 2018 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 1988-1996.

[15] MIAO C L, SU L, JIANG W J, et al. A lightweight privacy-preserving truth discovery framework for mobile crowd sensing systems[C]//Proceedings of IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.

[16] YANG Y, LIU X M, DENG R H, et al. Lightweight sharable and traceable secure mobile health system[J]. IEEE Transactions on Dependable and Secure Computing, 2020, 17(1): 78-91.

[17] ZHANG C, ZHU L H, XU C, et al. Reliable and privacy-preserving truth discovery for mobile crowdsensing sys-

tems[J]. IEEE Transactions on Dependable and Secure Computing, 2019(99): 1.

[18] ZHI S, YANG F, ZHU Z Y, et al. Dynamic truth discovery on numerical data[C]//Proceedings of 2018 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2018: 817-826.

[作者简介]

何英杰（1987− ），男，中国电子科技集团公司第二十二研究所高级工程师，主要研究方向为物联网通信技术。

李启伟（1985− ），男，中国电子科技集团公司第二十二研究所工程师，主要研究方向为通信技术和物联网通信技术。

孙涵（1992− ），女，中国电子科技集团公司第二十二研究所工程师，主要研究方向为物联网。

郜迪（1993− ），男，中国电子科技集团公司第二十二研究所工程师，主要研究方向为物联网嵌入式技术。

董剑峰（1993− ），男，中国电子科技集团公司第二十二研究所工程师，主要研究方向为物联网通信技术。

杨书华（1963− ），男，中国电子科技集团公司第二十二研究所研究员，主要研究方向为嵌入式系统和物联网技术。